TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

VOLUME 22
1921

PUBLISHED BY THE SOCIETY
LANCASTER, PA., AND NEW YORK
1921

1

ARITHMETICAL PARAPHRASES*

BY

E. T. BELL

I. Introduction

1. Let

$$\xi_i \equiv (x_{i1}, x_{i2}, \cdots, x_{ia_i}) \qquad (i = 1, 2, \cdots, r),$$

$$\eta_j \equiv (y_{j1}, y_{j2}, \cdots, y_{jb_j}) \qquad (j = 1, 2, \cdots, s),$$

denote $(r + s)$ one-rowed matrices of independent variables, no pair of matrices having a variable in common. Write

$$-\xi_i \equiv (-x_{i1}, -x_{i2}, \cdots, -x_{ia_i}) \equiv -(x_{i1}, x_{i2}, \cdots, x_{ia_i}),$$

and similarly for $-\eta_j$. If in η_j each $y = 0$, η_j is said to vanish. Let

$$(1) \qquad f(\xi_1, \xi_2, \cdots, \xi_r | \eta_1, \eta_2, \cdots, \eta_s)$$

denote a function which exists and has a determinate value for all integral values $\gtreqless 0$ of the x, y in ξ, η; which remains unchanged in value when any one of the ξ is replaced by its negative, and which changes sign and vanishes with each of the η. Similarly

$$(2) \qquad g(\xi_1, \xi_2, \cdots, \xi_r |), \qquad h(| \eta_1, \eta_2, \cdots, \eta_s)$$

exist and are determinate for all integral values of the x, y in ξ, η respectively; the value of g is unchanged when any one of the ξ is replaced by its negative; h changes sign and vanishes with each η.

It is emphasized, once for all, that beyond these restrictions f, g, h are wholly arbitrary.

As examples of the bar notation,

$$f(x, y|) = f(-x, y|) = f(x, -y|);$$

$$f(x|y) = f(-x|y) = -f(x|-y);$$

$$f(|x, y) = -f(|-x, y) = -f(|x, -y);$$

$$f((x, y)|z) = f((-x, -y)|z) = -f((x, y)|-z);$$

$$f((x, y, z), (u, w)|(t, v)) = f((-x, -y, -z), (u, w)|(t, v))$$

$$= f((x, y, z), (-u, -w)|(t, v)) = -f((x, y, z), (u, w)|(-t, -v)).$$

* Read before the San Francisco Section of the Society, October, 1918.

2. The parity of the f in (1) is denoted by

$$p(a_1, a_2, \cdots, a_r | b_1, b_2, \cdots, b_s); \tag{3}$$

and the respective parities of g, h in (2) are

$$p(a_1, a_2, \cdots, a_r | 0), \qquad p(0 | b_1, b_2, \cdots, b_s), \tag{4}$$

the notation being obvious. The positive integers

$$\omega = \sum_{i=1}^{r} a_i + \sum_{j=1}^{s} b_j, \qquad \delta = r + s, \tag{5}$$

are called the order and degree respectively of f. Similarly for g, h. When

$$a_i = 1 = b_j \quad (i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s),$$

the parities (3), (4) are written respectively:

$$p(1^r | 1^s), \ p(1^r | 0), \ p(0 | 1^s). \tag{6}$$

Likewise, if α_j of the a_i each $= a_j$, and β_i of the b_j each $= b_i$, the parities (3), (4) are written (the order of the a's or b's within $(|)$ is immaterial),

$$p(a_1^{\alpha_1}, a_2^{\alpha_2}, \cdots | b_1^{\beta_1}, b_2^{\beta_2}, \cdots); \ p(a_1^{\alpha_1}, a_2^{\alpha_2}, \cdots | 0); \ p(0 | b_1^{\beta_1}, b_2^{\beta_2}, \cdots). \tag{7}$$

From the definitions, an f whose parity is $p(1^r | 1^s)$ is a function of $(r + s)$ single independent variables, even separately in r of them, odd in each of the remaining s variables, and vanishing with each of the s. The corresponding statement for a function of parity $p(1^r | 0)$ follows on supposing $s = 0$; similarly for one of parity $p(0 | 1^s)$, on supposing $r = 0$. Henceforth we shall in general consider it unnecessary to give separate statements for f, g, h of (1), (2), regarding all as implicit in the statement for (1). The parity of a constant is considered $= p(0 | 0)$.

3. Without difficulty it may be shown* that an f of order ω and degree δ is

* For this result and that of § 4, cf. Bell, Bulletin of the American Mathematical Society, vol. 25 (1918–19), p. 313. The proofs follow readily from the fundamental identities (52), (53) of § 33, and (59), (60) of § 35. On account of its interest we add the following alternative proof. We are concerned in §§ 3, 4 with a generalization of the expression of a function as the sum of an odd and an even function. Thus

$$(A) \quad \begin{aligned} 2f(x) &= [f(x) + f(-x)] + [f(x) - f(-x)] \equiv \phi_0(x) + \phi_1(x), \\ 2f(-x) &= [f(x) + f(-x)] - [f(x) - f(-x)] \equiv \phi_0(x) - \phi_1(x). \end{aligned}$$

If now f is a general function of ω variables $x = x_1, \cdots, x_\omega$, then singling out x_1 we define $\phi_0(x)$, $\phi_1(x)$. In ϕ_0, ϕ_1 we single out x_2, and proceeding as in (A) obtain ϕ_{00}, ϕ_{01}, ϕ_{10}, ϕ_{11}, where 1, 0 indicates oddness or evenness respectively in the variables in order. Proceeding thus we have eventually 2^ω functions

$$\phi_{i_1 i_2 \ldots i_\omega} \qquad (i_j = 0, 1;\ j = 1, \cdots, \omega),$$

of parities $p(1^\alpha | 1^\beta)$, $(\alpha + \beta = \omega)$.

On the other hand we apply to $f(x_1, x_2, \cdots, x_\omega) = f_{000 \ldots 0}$ the operations of the group G

linearly expressible in terms of $2^{\omega-\delta}$ suitably chosen functions, all of whose parities are of the form $p(1^\alpha | 1^\beta)$, where $\alpha + \beta = \omega$.

4. Removing the restriction that the f in (1) shall vanish with each η, we get what we shall call a special f of parity (3). E.g., $wz/(x+y)$ is a special f of parity $p(0|1, 1, 2)$, $= p(0|1, 2, 1)$, etc. Clearly, parity has no relevance in regard to a perfectly arbitrary function of n variables; such a function is not necessarily even or odd in any one of its variables or in any matrix ξ, η of its variables. It is easy to show, however, that an arbitrary function of n variables is linearly expressible in terms of 2^n suitably chosen special f's, all of whose parities are of the form $p(1^\alpha | 1^\beta)$ where $\alpha + \beta = n$. This result and that of § 3 are basic in the subsequent discussion.

5. In addition to the functions already defined, we shall consider others, ϕ, having the same parities as f, g, h but further restricted, e.g., as to alterance, invariance under the substitutions of a finite group on the x, y, etc., the essential feature being change or invariance of sign under permutation of the variables. For a reason appearing presently, all functions f, g, h, ϕ are called L-functions, where the L stands for Liouville. Functions ϕ, and functions F, G, H, Φ which satisfy the same conditions of parity as f, g, h, ϕ but which also implicitly satisfy further conditions, as e.g., continuity, differentiability, etc., with respect to some or all of the x, y variables, are called restricted L-functions. The explicit restrictions on a given ϕ, which so far as this paper is concerned* are only of the nature that ϕ is unaltered to within sign under permutations of the variables, will be exhibited by stating the equations which express them. Thus,

$$\phi((x, y)|z) = -\phi((y, x)|z),$$

expresses that $\phi((x, y)|z)$, of parity $p(2|1)$, in addition to satisfying the parity equations

$$\phi((x, y)|z) = \phi((-x, -y)|z) = -\phi((x, y)|-z),$$

of order 2^ω of changes of sign of the variables and obtain 2^ω functions

$$f_{i_1 i_2 \dots i_\omega} \qquad (i_j = 0, 1),$$

where $i_j = 0, 1$ according as in $f = f_{000\dots 0}$ the sign of x_i has not or has been changed. *Then*, by repeated application of (A) we obtain a linear transformation with coefficients ± 1 which expresses the set of 2^ω functions $2^\omega f_{i_1 i_2 \dots i_\omega}$ in terms of the 2^ω functions ϕ. This is true therefore of the one function $2^\omega f_{000\dots 0} = 2^\omega f$.

If in particular (§ 3) f has degree δ, then f is unaltered to within sign by a subgroup (necessarily invariant since G is abelian) of G of order 2^δ. An operation such as (A) becomes the identity when f itself has parity, and the number of functions $f_{i_1 \dots i_\omega}$, $\phi_{i_1 \dots i_\omega}$, reduces to $2^{\omega-\delta}$ and the linear transformation between them contains the integer factor $2^{\omega-\delta}$.

* Other restrictions of great use in applications are of the kinds (i) $\phi(x|\,) = 1, 0$ according as x is or is not the $(2r-1)$th power of an integer; (ii) $\phi(x|\,) = 1, 0$ according as x is or is not divisible by a given integer, and a similar restriction upon $\phi(|x)$; (iii) the obvious extensions of these to ϕ's of several variables. Examples of these will be given in papers to appear elsewhere.

which are implicit in the bar notation, is alternating in x, y.

A set of equations expressing restrictions may imply further restrictions. For example we find theorems for restricted L-functions, ϕ, of order 4, the restrictions first presenting themselves in the form:*

$$\phi(x, y, z, w) = \phi(y, x, z, -w) = -\phi(x, -y, w, z).$$

From these we infer, among others:

$$\phi(x, y, z, w) = \phi(-x, -y, -z, -w) = -\phi(y, -x, -w, z).$$

Hence $\phi(x, y, z, w)$ may be represented by $\phi((x, y, z, w)|)$; and we have the canonical set of restrictions:

$$\phi((x, y, z, w)|) = \phi((y, x, z, -w)|) = -\phi((x, -y, w, z)|);$$

a set being canonical when it includes the parity conditions and a minimum number of restrictions from which all may be inferred.†

It will be shown, when we consider restrictions in detail, that a canonical set for a restricted L-function ϕ, of order ω, may always be found by determining the group to which a certain algebraic form on ω letters associated with ϕ belongs. This, at first sight, is rather remarkable, as the L-functions (cf. § 1), are not necessarily algebraic.‡

6. With ξ, η as in § 1, consider the implicitly restricted L-functions:

$$F(\xi_1, \xi_2, \cdots, \xi_r | \eta_1, \eta_2, \cdots, \eta_s)$$
$$\equiv \sum_{m=1}^{k} c_m \left[\prod_{i=1}^{r} \cos\left(\sum_{n=1}^{a_i} \alpha_{min}\, x_{in}\right) \cdot \prod_{j=1}^{s} \sin\left(\sum_{n=1}^{b_j} \beta_{mjn}\, y_{jn}\right)\right];$$

$$G(\xi_1, \xi_2, \cdots, \xi_r|) \equiv \sum_{m=1}^{k} c_m \left[\prod_{i=1}^{r} \cos\left(\sum_{n=1}^{a_i} \alpha_{min}\, x_{in}\right)\right];$$

$$H(|\eta_1, \eta_2, \cdots, \eta_s) \equiv \sum_{m=1}^{k} c_m \left[\prod_{j=1}^{s} \sin\left(\sum_{n=1}^{b_j} \beta_{mjn}\, y_{jn}\right)\right].$$

* An example occurs among the illustrations, § 15 (19a).

† The following alternative statement may be made. In the notation of § 3, footnote, the permutations of the variables under which f is unaltered to within sign generate with G an enlarged group Γ under which G is invariant to within sign. Thus a *canonical* set of restrictions may be described as one which gives the generators of G and a minimum number of generators of the factor group of G under Γ, i.e., the permutation group.

‡ Detailed consideration of this point having been omitted to save space, we shall give here sufficient indications of the course to be followed, from which the whole process can easily be reconstructed. The algebraic form mentioned is that which is deduced from the reduced invariant I_ϵ defined, Bulletin of the American Mathematical Society, vol. 26, p. 217, § 9, as follows: each k-ad (ibid., p. 212, § 2) is to be replaced by the restricted L-functions F, G or H of this paper, § 6, on the same variables; the algebraic form is then the coefficient in this result of the general term in the x, y variables when the entire I_ϵ is expanded in powers of these variables.

Write

$$A_{mi} \equiv (\alpha_{mi1}, \alpha_{mi2}, \cdots, \alpha_{mia_i}) \qquad (i = 1, 2, \cdots, r);$$

$$B_{mj} \equiv (\beta_{mj1}, \beta_{mj2}, \cdots, \beta_{mjb_j}) \qquad (j = 1, 2, \cdots, s);$$

and let the c, α, β denote integers. Then, in its general form, the principle of paraphrase which we shall use is:

(i) If for all values of the x, y,

$$F(\xi_1, \xi_2, \cdots, \xi_r | \eta_1, \eta_2, \cdots, \eta_s) = 0, \tag{8}$$

then

$$\sum_{m=1}^{k} c_m f(A_{m1}, A_{m2}, \cdots, A_{mr} | B_{m1}, B_{m2}, \cdots, B_{ms}) = 0. \tag{8a}$$

(ii) If for all values of the x,

$$G(\xi_1, \xi_2, \cdots, \xi_r |) = 0, \tag{9}$$

then

$$\sum_{m=1}^{k} c_m g(A_{m1}, A_{m2}, \cdots, A_{mr} |) = 0. \tag{9a}$$

(iii) If for all values of the y,

$$H(| \eta_1, \eta_2, \cdots, \eta_s) = 0, \tag{10}$$

then

$$\sum_{m=1}^{k} c_m h(| B_{m1}, B_{m2}, \cdots, B_{ms}) = 0. \tag{10a}$$

In (8*a*), (9*a*), (10*a*), f, g, h are general L-functions as defined in § 1; and the principle asserts that the sine-cosine identities (8), (9), (10) may be paraphrased directly into (8*a*), (9*a*), (10*a*) respectively. By means of this simple principle, which we shall prove as required (cf. § 18 et seq.), the applications of the elliptic, hyperelliptic and theta functions to the theory of numbers are greatly extended. For, from the theories of these functions we write down identities (8), (9), (10) in which the A_{mi}, B_{mj} are matrices whose elements are linear functions of the divisors of integers belonging to certain linear or quadratic forms (more specifically defined in §§ 7, 8). The (8*a*), (9*a*), (10*a*) written down from the (8), (9), (10) then give, for special choices of the L-functions, as for example

$$f(x, y|) = y^{2n} + x^{2k} \cos \pi y, \qquad f(x, y|z) = (-1)^x y^{2k} \sin (z^3\pi/2),$$

$$f(|x) = x^{2n} \sin \frac{\pi}{x},$$

an inexhaustible source of arithmetical theorems. It will be noted that this principle effects the passage from circular to L-functions immediately without further analysis or transformations. Finally, it will be shown,* from a para-

* Cf. §§ 32–34.

phrase concerning L-functions of parity $p(a|0)$ that we can at once infer paraphrases in which the L-functions are of either of the parities $p(a_1, a_2|0)$, $p(0|a_1, a_2)$, where $a_1 + a_2 = a$. Similarly, from a paraphrase for L-functions of parity $p(0|b)$ follow immediately paraphrases for L-functions of parity $p(b_1|b_2)$, where $b_1 + b_2 = b$. Now obviously an L-function of parity $p(a_1, a_2, \cdots, a_r|b_1, b_2, \cdots, b_s)$ may be regarded as an L-function of any of the parities $p(a_i|0)$, $p(0|b_j)$, $(i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s)$. Applying the foregoing inferences successively to some or all of the a_i, b_j, we find that a paraphrase in which the L-functions are unrestricted of parity $p(a_1, a_2, \cdots, a_r|b_1, b_2, \cdots, b_s)$, degree δ, order ω, implies further paraphrases for unrestricted L-functions of order ω, and degree δ', where

$$\delta < \delta' \leqq \omega.$$

From the paraphrases for the functions of degree δ' may be readily built up paraphrases* for L-functions of order ω subject to restrictions as outlined in § 5.

7. Before illustrating the nature of the paraphrases we shall define the sense in which *separation* is used constantly throughout. Unless the contrary is explicitly stated, all integers now considered are positive and different from zero. Adopting Glaisher's convenient notation,† we use letters m to denote odd integers, letters n to denote arbitrary integers; and in reference to separations, m, n shall always, without further specification, have this significance. Letters d, δ denote positive integral divisors. Hence in $m = d\delta$ both d, δ are odd; in $n = d\delta$ either or both d, δ may be odd or even; and $n = 2^\alpha m$, in which $\alpha \geqq 0$, indicates the highest power of 2 that divides n. We shall be frequently concerned with three types of division, T_1, T_2, T_3:

$$T_1 : m = d\delta; \qquad T_2 : n = 2^\alpha m,\ m = d\delta; \qquad T_3 : n = d\delta. \tag{11}$$

Let $n, c, c_1, c_2, \cdots, c_r, c'_1, c'_2, \cdots, c'_s$ denote fixed integers, $n, c > 0$, the rest $\geqq 0$; $n_1, n_2, \cdots, n_r, n'_1, n'_2, \cdots, n'_s$ variable integers. Then, a separation of cn is the totality, $[S]$, of all solutions, $(2^\alpha d,\ \delta,\ 2^\alpha d_1,\ \delta_1,\ \cdots,\ n'_1,\ n'_2,\ \cdots)$, of such a system as

$$\begin{aligned} &cn = c_1 n_1 + c_2 n_2 + \cdots + c_r n_r + c'_1 n'^2_1 + c'_2 n'^2_2 + \cdots + c'_s n'^2_s, \\ &n = 2^\alpha m,\ n_1 = 2^{\alpha_1} m_1,\ \cdots,\ n_r = 2^{\alpha_r} m_r, \\ &n'_1 \gtreqless 0,\ n'_2 \geqq 0,\ \cdots,\ n'_s \gtrless 0, \\ &m = d\delta,\ m_1 = d_1 \delta_1,\ \cdots,\ m_r = d_r \delta_r, \\ &\alpha > 0,\ \alpha_1 \geqq 0,\ \cdots,\ \alpha_r = 0, \end{aligned} \tag{12}$$

whose essential characteristics are:

---

* The process is illustrated in Bulletin of the American Mathematical Society, vol. 26 (1919–20), p. 218, § 10, and elsewhere in the same paper.

† Kronecker used a similar notation in his memoirs on class-number relations; cf. Journal für Mathematik, vol. 57 (1860), p. 248.

(i) T_j $(j = 1, 2, 3)$ is given for each of $n_1, n_2, \cdots, n_r$;

(ii) the range of permissible values for each of the $n_1', n_2', \cdots, n_s'$ is specified, when it is other than $+1$ to $+\infty$; viz., the range, which may be any of $\gtreqless 0$, > 0, $\gtrsim 0$ according to the case, of permissible values for each of the $n_1', n_2', \cdots, n_s'$ is specified in a given separation. Similarly for the α's, which may range > 0, $\geqq 0$. The actual set given in (12) is merely a specimen separation. Thus $n_1' \geqq 0$, $n_2' > 0$, $n_3' \gtreqless 0$, $\alpha \geqq 0$, $\alpha_1 > 0$, $\alpha_2 > 0$, $\alpha_3 \geqq 0$ characterizes one definite separation; $n_1' > 0$, $n_2' \gtreqless 0$, $n_3' > 0$, $\alpha > 0$, $\alpha_1 \geqq 0$, $\alpha_2 \geqq 0$ characterizes another.

(iii) The coefficients c, c_i, c_j' are all positive.

When further conditions, e.g., $\delta_1 < \sqrt{m_1}$, are imposed, the separation is said to be restricted. The degree* of $[S]$ is the number of non-vanishing c_i, c_j'.

8. Let the degree of $[S]$ be ν; and denote by (S) a particular solution of (12):

$$(S) \equiv (\lambda_1, \lambda_2, \cdots, \lambda_\nu).$$

Form ω linear functions of the λ's:

$$\Lambda_i \equiv l_{i1}\lambda_1 + l_{i2}\lambda_2 + \cdots + l_{i\nu}\lambda_\nu \qquad (i = 1, 2, \cdots, \omega);$$

and denote by $F(z_1, z_2, \cdots, z_\omega)$ any L-function of order ω. Construct $F(\Lambda_1, \Lambda_2, \cdots, \Lambda_\omega)$ for each (S) in $[S]$. Since the c_i, $c_j' \geqq 0$, there will be only a finite number, k, of such F's; say

$$F(S_1),\ F(S_2),\ \cdots,\ F(S_k).$$

We shall be concerned with sums

$$\sum_{i=1}^{k} a_i F(S_i), \tag{13}$$

where the a_i denote constant integers, for L-functions of specified parities; and (13) is defined to be the integration of $a_i F(X_1, X_2, \cdots, X_\omega)$ over $[S]$, where

$$X_i \equiv l_{i1}x_1 + l_{i2}x_2 + \cdots + l_{i\nu}x_\nu \qquad (i = 1, 2, \cdots, \omega).$$

9. Separations are segregated into two main classes: linear, when $c_1' = c_2' = \cdots = c_s' = 0$; quadratic, when at least one $c_j' > 0$. Linear separations are further classified according to the types T_1, T_2, T_3; and quadratic, in addition to the specification of types for the η_i, according to the evenness or oddness

* The degree of $[S]$ expresses, as will be evident from the derivations of the paraphrases in Part II, section V, the greatest number of elliptic and theta series which are multiplied together in an identity furnishing L-function paraphrases whose integrations (§ 8) are over $[S]$. This has proved a useful clue in tracing certain of Liouville's more abstruse results to their elliptic-theta equivalents, cf. § 13.

of the n_i'. This classification is basic in connection with the subsequent classification and interlacing of the paraphrases, the latter depending naturally upon the former. The elliptic and theta series which we shall use are similarly classified before paraphrasing.

10. Paraphrases, which will be of the general form $\sum_{i=1}^{i=k} a_i F(S_i) = 0$, (cf. § 8), will be stated by giving the separations and corresponding integrations, which always, as in § 8, are with respect to the separations. For simplicity in writing, the L-functions under the $\sum$ will sometimes be indicated as follows:

$$f(x_1, x_2, \cdots, x_r, y_1, y_2, \cdots, y_s) \equiv F(x_1, x_2, \cdots, x_r | y_1, y_2, \cdots, y_s),$$

and the paraphrase written $\sum f() = 0$. Paraphrases in which the integrations are over several separations will be similarly written, the several separations being given separately by different systems of letters, thus:

$$n = m_1 + 2m_2; \qquad n = 2^{\alpha'} m' + m''; \qquad \cdots$$

$$m_1 = d_1 \delta_1, \qquad m_2 = d_2 \delta_2; \qquad m' = d' \delta', \qquad m'' = d'' \delta''; \qquad \cdots.$$

Always, unless it is explicitly given that they are restricted, the L-functions are general as defined in § 1.

11. To illustrate the concepts of this introduction we shall now give without proof* a few simple examples. These indicate the nature of the general formulas into which we later paraphrase certain parts of the theories of elliptic and theta functions. *References are at the end of this paper.*

As a first example we consider the following in detail. By a simple transformation it is easily shown to be identical with Liouville's 5, (f).

$$n = n' + n''; \qquad n = d\delta, \qquad n' = d'\delta', \qquad n'' = d''\delta'':$$

$$\begin{aligned}(14)\quad &\sum [f(d' - d'', \delta' + \delta''|) - f(d' + d'', \delta' - \delta''|)] \\ &= \sum [(d-1)\{f(0, d|) - f(d, 0|)\} + 2\sum_{r=1}^{d-1} \{f(\delta, r|) - f(r, \delta|)\}].\end{aligned}$$

Here an L-function of parity $p(1^2|0)$ is integrated over a linear separation of degree 2, and of a type that may be conveniently designated by T_3^2. The precise nature of (14) will be evident from a numerical example. Let $n = 5$; then:

* The first example is proved in part II, § 23. The paraphrase of the $t(x, y)$ identity in § 14 is immediate from the series for the doubly periodic functions of the second kind given in Part II, § 16; that of (ii) is a translation of the trigonometric identity obtained on equating coefficients of q^m, the series for the functions being written down from those given by G. Humbert in Journal de Mathématiques pures et appliquées, (6) 3, vol. 72 (1907), p. 350, first formula in (5), and from Hermite, *Œuvres*, vol. 2, p. 244, formula 1.

$(n', n'') = (1, 4), (2, 3), (3, 2), (4, 1)$:

$$\begin{array}{l|l|l|l|l|}
(n', n'') = & (1,4) & (2,3) & (3,2) & (4,1) \\
(d', \delta') = & (1,1) & (1,2),(2,1) & (1,3),(3,1) & (1,4),(2,2),(4,1) \\
(d'', \delta'') = & (1,4),(2,2),(4,1) & (1,3),(3,1) & (1,2),(2,1) & (1,1)
\end{array},$$

whence, for the successive (n', n'') the values of $(d' \mp d'', \delta' \pm \delta'')$ are

$$\begin{array}{c|l|l|}
(n', n'') & (d' - d'', \delta' + \delta'') & (d' + d'', \delta' - \delta'') \\
(1,4) & (0,5),(-1,3),(-3,2) & (2,-3),(3,-1),(5,0) \\
(2,3) & (0,5),(-2,3),(1,4),(-1,2) & (2,-1),(4,1),(3,-2),(5,0) \\
(3,2) & (0,5),(-1,4),(2,3),(1,2) & (2,1),(3,2),(4,-1),(5,0) \\
(4,1) & (0,5),(1,3),(3,2) & (2,3),(3,1),(5,0)
\end{array}.$$

Since $f(x, y|) = f(-x, y|) = f(x, -y|)$, we have, on writing

$$f(x, y) \equiv f(x, y|):$$

$$[4f(0,5) + 2f(1,3) + 2f(3,2) + 2f(2,3) + 2f(1,4) + 2f(1,2)]$$
$$- [4f(5,0) + 2f(3,1) + 2f(3,2) + 2f(2,3) + 2f(4,1) + 2f(2,1)],$$

for the left of (14); and this reduces to:

$$4[f(0,5) - f(5,0)]$$
$$+ 2[f(1,2) - f(2,1) + f(1,3) - f(3,1) + f(1,4) - f(4,1)].$$

For $n = 5$, we have $(d, \delta) = (1, 5), (5, 1)$; and the right of (14) is:

$$(1-1)\{f(0,1) - (1,0)\} + (5-1)\{f(0,5) - f(5,0)\}$$
$$+ 2\sum_{r=1}^{4} \{f(1,r) - f(r,1)\},$$

which agrees with the value found for the left.

12. All of Liouville's formulas for functions whose order or degree exceeds unity have in common one feature which is truly remarkable. To see it for (14), an inspection of the numerical example will show that in each $f(d' \mp d'', \delta' \pm \delta''|)$, d', d'' are associated with *their own* conjugates δ', δ''. That is, if all the resolutions of n in the form $n = d'\delta' + d''\delta''$ are

$$n = d_1'\delta_1' + d_1''\delta_1'' = d_2'\delta_2' + d_2''\delta_2'' = \cdots = d_k'\delta_k' + d_k''\delta_k'',$$

the left of (14) is

$$\text{(14a)} \qquad \sum_{i=1}^{k}[f(d_i' - d_i'', \delta_i' + \delta_i''|) - f(d_i' + d_i'', \delta_i' - \delta_i''|)],$$

and not (for instance) what the single $\sum$ notation might equally well be used to express:

$$\text{(14b)} \qquad \sum_{i=1}^{k}\sum_{j=1}^{k}[f(d_i' - d_i'', \delta_j' + \delta_j''|) - f(d_i' + d_i'', \delta_j' - \delta_j''|)].$$

Wherever in the sequel d, δ, d', δ', $\cdots$ are associated together in an L-function, the d, δ, the d', δ', $\cdots$ are conjugates; and the $\sum$ has the meaning of (14*a*), never of (14*b*). When we come to examine the elliptic and theta series for paraphrases, we shall see that paraphrases involving sums of the kind (14*b*) may be written down with great ease, while those of the Liouville kind, in which the sums are of the form (14*a*) while also readily deducible from certain of the expansions, are much less common, and therefore of correspondingly greater interest. The applications of the (14*a*) kind seem also to be of more importance than those of the (14*b*). It is interesting to note that paraphrases for sums of L-functions of degrees or orders > 1, in which the divisors are associated with their own conjugates as arguments of the L-functions, are implicit in Jacobi's memoirs on rotation, also in many of Hermite's earlier (and some of his later) papers on elliptic functions,* but not in the *Fundamenta Nova*. Nor do they occur in Schwarz' '*Sammlung,*' although many of the lists in that work may be prepared easily in a form suitable for the deduction of such paraphrases. A few of Kronecker's uncollected notes on elliptic function series also contain developments leading to (14*a*) paraphrases.

13. Passing to a more significant illustration of (14), we choose for $f(x, y|)$ the (implicitly) restricted L-function $\cos 2xu \cos 2yv$ in which u, v are parameters. After some simple reductions, (14) becomes:

$$\begin{aligned} &2\sum \sin 2(d'u + \delta'v)\sin 2(d''u - \delta''v) \\ \text{(15)} \qquad &= \sum d(\cos 2du - \cos 2dv) \\ &+ \sum(\cot v \cos 2du \sin 2\delta v - \cot u \sin 2du \cos 2\delta v), \end{aligned}$$

* References to which are given in Part II where the series are considered.

which is the result of equating coefficients of $q^{n/2}$ in:

$$(16)\qquad \frac{\vartheta_1'\,\vartheta_1(u+v)}{\vartheta_1(u)\,\vartheta_1(v)}\cdot\frac{\vartheta_1'\,\vartheta_1(u-v)}{\vartheta_1(u)\,\vartheta_1(-v)}\equiv\left[\vartheta_2\,\vartheta_3\frac{\vartheta_0(u)}{\vartheta_1(u)}\right]^2-\left[\vartheta_2\,\vartheta_3\frac{\vartheta_0(v)}{\vartheta_1(v)}\right]^2.$$

In paraphrasing these steps are reversed. We start with (16), deduce (15), change (15) by separating trigonometric products into sums to the form (9), and paraphrase the result by (9a) immediately into (14). We note that, (15) being a very special case of (14); and (16), when considered merely as an identity between series, being deducible from (15) by a simple reversal of the steps which lead from (16) to (15), in a sense (14) includes (16) as a special case. There is, however, nothing in (14) that gives any immediate information concerning the periodicity, pseudo or real, of the quotients in (16). From this point of view, (16) is more general than (14). Against this may be put the following remarks of Liouville, which accord with the first view: "En effet mes formules se rattachent aussi à la théorie des fonctions elliptiques, seulement elles contiennent plutôt cette théorie qu'elles n'en dependent. . . . On n'a pas plus peine à y arriver au moyen des fonctions elliptiques.* Il y a là un genre de traduction que l'habitude rend facile" (19; p. 44). Again, (speaking of his general formulas): "Elles donnent naissance à des équations entre des séries qui contiennent comme cas particulier celles de la théorie des fonctions elliptiques" (19; p. 41).

From the present standpoint, (14), (15) are abstractly identical; but (14), as shown by numerous applications made of it by Liouville and others, presents the arithmetical information implicit in (15) or (16) in the more suggestive and usable form.

14. The diversity of the paraphrases is evident from the two following, selected at random from those found systematically in the sequel. Each is but one of several interpretations of the corresponding theta formula from which it is deduced.

(i) Write $t(x,y)\equiv\vartheta_1'\,\vartheta_1(x+y)/\vartheta_0(x)\,\vartheta_0(y)$, and denote by $t_w(x,y)$ the w-derivative of $t(x,y)$. Then,

$$t_x(x,y)-t_y(x,y)=t(x,y)\left[\frac{\vartheta_0'(y)}{\vartheta_0(y)}-\frac{\vartheta_0'(x)}{\vartheta_0(x)}\right],$$

which paraphrases into the elegant result:

$$m=m_1+2n_2;\quad n_2=2^{\alpha_2}m_2;\quad m=d\delta,\quad m_1=d_1\,\delta_1,\quad m_2=d_2\,\delta_2:$$

* The process of proof which Liouville suggests for the deduction from elliptic functions of his paraphrases concerning L-functions of order 1 cannot be extended to deduce paraphrases in which the order exceeds 1. Hence it will not be followed here. Again, regarding its proposed application to the functions of order 1, Liouville's method assumes that the functions are expansible in a Fourier series, an assumption which would not be justified for L-functions as defined in § 1. Liouville does not indicate from what elliptic function identities his theorems may be deduced.

$$(17)\quad 4\sum[\Phi(\delta_1, d_1 - 2^{a_2+1} d_2) + \Phi(d_1 + 2^{a_2+1} d_2, \delta_1)]$$
$$= \sum(d - \delta)\Phi(d, \delta),$$

where Φ is any one of the restricted L-functions, ϕ, ψ, χ, defined by:

$$\phi((x, y)|) = -\phi((y, x)|); \qquad \psi(x, y|) = -\psi(y, x|);$$
$$\chi(|x, y) = -\chi(|y, x).$$

The respective parities of ϕ, ψ, χ are $p(2|0)$, $p(1^2|0)$, $p(0|1^2)$: and the functions are (explicitly) restricted because subject to one other condition, here change of sign with interchange of variables, in addition to those of parity. Illustrative of general processes considered in §§ 25, 32, 36, the paraphrase for ϕ implies both the ψ and the χ paraphrases, which are independent; and from ψ, χ together, it is easy to infer ϕ. Special cases of interest arise for the choices, obviously legitimate:

$$\phi((x, y)|) = f((x, y)|) - f((y, x)|);$$
$$\psi(x, y|) = f(x, y|) - f(y, x|);$$
$$\chi(|x, y) = f(|x, y) - f(|y, x).$$

In fact, the Φ-paraphrase first presents itself for this ϕ; and by the processes cited, the ϕ-paraphrase may be at once replaced by the Φ-form.

(ii) One paraphrase of the identity

$$\vartheta_2 \frac{\vartheta_1(x)\vartheta_2(x)}{\vartheta_0(x)} \cdot \vartheta_2\vartheta_3 \frac{\vartheta_1(x)}{\vartheta_0(x)} = \vartheta_2^2\vartheta_3 \frac{\vartheta_1^2(x)\vartheta_2(x)}{\vartheta_0^2(x)}$$

is for a restricted linear separation of degree 2, and a function of parity $p(1|0)$:

$$m = l_1 + 2m_2 = d\delta \equiv 1 \mod 4; \qquad m_1 \lesseqgtr \sqrt{m}; \qquad m_2 = d_2\,\delta_2;$$
$$l_1 = d_1\,\delta_1 \equiv -1 \mod 4; \qquad d_1 > \sqrt{l_1}:$$

$$(18)\quad \sum\left[f\left(\frac{d_1 + \delta_1}{2} - d_2\Big|\right) - f\left(\frac{d_1 + \delta_1}{2} + d_2\Big|\right)\right]$$
$$= \sum\left[F(m - m_1^2)f(m_1|) - \left(\frac{d - \delta}{2}\right)f\left(\frac{d + \delta}{2}\Big|\right)\right],$$

where $F(n)$ is, with the usual conventions, the number of uneven classes, for the determinant $-n$, of binary quadratic forms. Such formulas in which the L-functions are of orders and degrees > 1, containing those in which the order or degree is 1 as special cases, may be derived with great ease on combining the series in Part II, § 15, with those given by Humbert (loc. cit.), and form the subject of a separate paper. For the L-functions suitably specialized

these formulas give, among others, the class number formulas of Kronecker, Hermite, Liouville and others.

15. To have an illustration of the processes considered in §§ 35, 36, we transcribe the following.

$$m = m_1^2 + 8n_2; \qquad n_2 = d_2\,\delta_2:$$

$$\text{(19)}\quad 2\sum(-1)^{(m_1+1)/2}\phi(2d_2-m_1, 2\delta_2+m_1|) = \epsilon(m)(-1)^{(\sqrt{m}-1)/2}\Big[\phi(1, \sqrt{m}|) + \sum_{r=1}^{(\sqrt{m}-1)/2}\{\phi(2r-1, \sqrt{m}|) - \phi(\sqrt{m}, 2r-1|)\}\Big],$$

where $\epsilon(n) = 1$ or 0 according as n is or is not a square; and $\phi(x, y|)$ is subject to the restriction $\phi(x, y|) = -\phi(y, x|)$. The separation here is quadratic and unrestricted of degree 2. For the same separation, we find by a process of linear transformation of the variables in (19), the following transform of it:

$$\text{(19a)}\quad \begin{aligned} &2\sum(-1)^{(m_1+1)/2}\phi((d_2+\delta_2, d_2-\delta_2-m_1, 2d_2-m_1, 2\delta_2+m_1)|) \\ &= \epsilon(m)(-1)^{(\sqrt{m}-1)/2}\Big[\phi\Big(\Big(\frac{1+\sqrt{m}}{2}, \frac{1-\sqrt{m}}{2}, 1, \sqrt{m}\Big)\Big|\Big) \\ &+ \sum_{r=1}^{(\sqrt{m}-1)/2}\Big\{\phi\Big(\Big(r+\frac{1+\sqrt{m}}{2}, r+\frac{1-\sqrt{m}}{2}, 2r+1, \sqrt{m}\Big)\Big|\Big) \\ &+ \phi\Big(\Big(r-\frac{1+\sqrt{m}}{2}, r-\frac{1-\sqrt{m}}{2}, 2r-1, \sqrt{m}\Big)\Big|\Big)\Big\}\Big], \end{aligned}$$

where ϕ is subject to the restrictions, forming a canonical set (§ 5):

$$\phi((x, y, z, w)|) = \phi((y, x, z, -w)|) = -\phi((x, -y, w, z)|).$$

The transformation for passing from (19) to (19a) is briefly indicated in § 35 (end). It is a good exercise in the bar notation to verify (19), (19a) for $m = 17, 25$.

For the same system of arguments, $d_2 + \delta_2$, etc., as in (19a), the linear transformation converting (19) into (19a) gives also 15 more paraphrases, seven of which are for restricted functions, and eight for unrestricted. This indicates the fertility of the method. These paraphrases, together with an infinity more, are all consequences of the obvious identity:

$$\vartheta_1(x-y)\cdot\frac{\vartheta_1'\,\vartheta_1(x+y)}{\vartheta_1(x)\,\vartheta_1(y)} + \vartheta_1(x+y)\cdot\frac{\vartheta_1'\,\vartheta_1(x-y)}{\vartheta_1(x)\,\vartheta_1(-y)} = 0.$$

From this identity, when $\vartheta_1'\,\vartheta_1(x \pm y)/\vartheta_1(x)\,\vartheta_1(\pm y)$ are replaced by their Fourier expansions given in Part II (or written out independently in the usual way), and § 36, the origin of the restriction imposed upon $\phi(x, y|)$ in (19), is sufficiently evident.

We may mention here some general results which form part of a later investigation. The example just given illustrates the concept of a class of paraphrases; two paraphrases being equivalent when either may be transformed into the other by a linear transformation of the variables, the coefficients of the transformation being rational. All paraphrases equivalent to one another constitute a class. In each class there is one and only one subclass, the reduced class, such that the order of the functions in any member of the class cannot be further reduced by linear transformations on the variables, and such that any member of the class may be transformed into any other by a transformation with coefficients ± 1 on the variables. The reduced class is said to be represented by any one of its members. In the above, (19) represents a reduced class; and (19*a*) is equivalent to (19). It is easily seen from §§ 31, 32 that (19*a*) includes (19) as a special case; but it is less obvious that (19) includes (19*a*).

16. Proofs for most of Liouville's general formulas will be found in the cited papers of Smith, Pepin, Mathews and Meissner. All of these use the method of Dirichlet in modified or extended form, to which Liouville himself repeatedly refers; but this method (cf. Bachmann, p. 366), offers no suggestion either as to proper assumptions to be made regarding the parity or restrictions of the functions, or to the constitution of the separations for a given function. It is, in fact, a process of à posteriori verification. By the method of paraphrase the questions concerning the nature of the functions and separations receive immediate answers on an examination of the class of series from which the paraphrases are derived. As it has been suggested by Bachmann (p. 433) that the source of Liouville's theorems was a consideration of the transformation of bilinear forms on four variables (as given, for example by Kronecker, *Werke*, vol. 1, p. 143), we shall state what seem the principal advantages of deriving them as paraphrases primarily of the elliptic-theta identities. Considering, for example, (16), it may be made, by simple algebraic or analytical transformations, to yield many more paraphrases in addition to (14), some of which are for quadratic separations, some for separations of degrees $3, 4$, and others for restricted or unrestricted functions of orders $1, 2, 3, 4, \cdots$ integrated over linear separations of degree 2. Even with the end-results before us, it is a matter of considerable difficulty to transform these into each other by Dirichlet's method as used (in amplified form) by Smith, Pepin, Meissner and others; and this method would seem to be the natural modification of Kronecker's transformation processes to be used for this purpose. But the most important advantage is that we have in the method of paraphrase what that of Dirichlet has not yet given, a direct and powerful means for the discovery of new paraphrases, which severally, as Bachmann says of this class of theorems (l. c., p. 366), " eine schier unerschöpfliche Fundgrube für zahlen-

theoretische Sätze darbieten." We shall not derive all of Liouville's general formulas en bloc by the method of paraphrase, although this may easily be done if desired, but shall derive them incidentally as they arise in applying the following developments to the elliptic and theta series.

17. An inspection of the numerical example in § 11, reveals the important fact, otherwise obvious from §§ 1, 6, that (14) is ultimately an identity between sets of absolute values of integers; two sets, $(|a_1|, |a_2|)$, $(|b_1|, |b_2|)$, being identical when and only when $|a_1| = |b_1|$ and $|a_2| = |b_2|$. The like, considerably generalized, will be evident for functions of parity

$$p(a_1, a_2, \cdots, a_r | b_1, b_2, \cdots, b_s).$$

Hence we next examine the properties of sets of matrices of absolute values. On them we shall base a proof, by new but simple considerations, of the legitimacy of the paraphrase process outlined in § 6, in sufficient detail to derive all the paraphrases first arising in the theory of the elliptic and theta functions. The process for functions of parity

$$p(a_1, a_2, \cdots, a_r | b_1, b_2, \cdots, b_s)$$

will appear as a corollary of that for functions of parity $p(a_1, a_2, \cdots, a_r | 0)$, and the latter as a corollary of the process for $p(a_1 | 0)$, which in turn follows from that for $p(1 | 0)$.

II. Sets of matrices and *L*-functions

18. The equality between matrices, $(a_1, a_2, \cdots, a_r) = (a_1', a_2', \cdots, a_s')$, implies $s = r$ and $a_i = a_i'$, $(i = 1, \cdots, r)$. If $a_i = 0$, $(i = 1, \cdots, r)$, the matrix is the zero matrix, $(0, 0, \cdots, 0)$, $\equiv (0)_r$. A set is a collection of things independently of their order. We shall write the matrix of absolute values

$$(|x_{i1}|, |x_{i2}|, \cdots, |x_{ir}|) \equiv (|x_i|)_r;$$

and the set of $(n - j)$ matrices

$$(|x_{j+1}|)_r, (|x_{j+2}|)_r, \cdots, (|x_n|)_r \qquad (n \geqq j + 1, j \geqq 0),$$

will be denoted by either of

$$\int_j^n (|x_i|)_r, \int_j^n (|x_{i1}|, |x_{i2}|, \cdots, |x_{ir}|);$$

and when all the $(n - j)$ matrices are zero, the set will be written, as convenient, in any of the forms

$$\int_j^n (0)_r, \int_j^n (0, 0, \cdots, 0), (n - j) \int (0)_r, (n - j) \int (0, 0, \cdots, 0),$$

the 0 in $(0, 0, \cdots, 0)$ being repeated r times. Two sets are equal:

$$\int_j^n (|x_i|)_r = \int_{j'}^{n'} (|x'_i|)_{r'} \tag{20}$$

when and only when the $(|x_i|)_r$ are a permutation of the $(|x'_i|)_{r'}$; and hence in particular only when $r' = r$ and $n' - j' = n - j$.

19. The sum (logical sum) of two sets is that set which consists of all the matrices in either set. Hence addition of sets is commutative and associative, and

$$\int_j^\lambda (|x_i|)_r + \int_\lambda^n (|x_i|)_r = \int_j^n (|x_i|)_r \qquad (0 \leqq j < \lambda < n). \tag{21}$$

20. An obvious property of sets for which we shall have frequent use is that the same $|\alpha|$ may be inserted in homologous places of equal sets without destroying their equality, viz., (20) implies

$$\begin{aligned}\int_j^n (|x_{i1}|, \cdots, |x_{is-1}|, |\alpha|, |x_{is}|, \cdots, |x_{ir}|)& \\ = \int_{j'}^{n'} (|x'_{i1}|, \cdots, |x'_{is-1}|, |\alpha|, |x'_{is}|, \cdots, |x'_{ir}|)&.\end{aligned} \tag{22}$$

Again, from the definitions, if $p, q, \cdots, t$ are any of the integers $1, 2, \cdots, r$, (20) implies

$$\int_j^n (|x_{ip}|, |x_{iq}|, \cdots, |x_{it}|) = \int_{j'}^{n'} (|x'_{ip}|, |x'_{iq}|, \cdots, |x'_{it}|). \tag{23}$$

21. Immediately from the definitions of §§ 1, 18:

Lemma 1. *If* $\int_0^n (|x_i|)_r = \int_0^n (|x'_i|)_r$, *then*

$$\sum_{i=1}^n f(x_{i1}, x_{i2}, \cdots, x_{ir}|) = \sum_{i=1}^n f(x'_{i1}, x'_{i2}, \cdots, x'_{ir}|).$$

In the same way, or as an obvious corollary:

$$\begin{aligned}\int_0^n (|x_i|)_r + \int_0^t (|y_i|)_r + \cdots + \int_0^p (|z_i|)_r& \\ = \int_0^n (|x'_i|)_r + \int_0^t (|y'_i|)_r + \cdots + \int_0^p (|z'_i|)_r&\end{aligned}$$

implies

$$\begin{aligned}\sum_{i=1}^n f(x_{i1}, x_{i2}, \cdots, x_{ir}|) + \sum_{i=1}^t f(y_{i1}, y_{i2}, \cdots, y_{ir}|)& \\ + \cdots + \sum_{i=1}^p f(z_{i1}, z_{i2}, \cdots, z_{ir}|)&\end{aligned}$$

$$= \sum_{i=1}^{n} f(x'_{i1}, x'_{i2}, \cdots, x'_{ir}|) + \sum_{i=1}^{t} f(y'_{i1}, y'_{i2}, \cdots, y'_{ir}|) + \cdots + \sum_{i=1}^{p} f(z'_{i1}, z'_{i2}, \cdots, z'_{ir}|).$$

22. For the passage from circular to L-functions the following lemmas* are fundamental.

LEMMA 2. *If the a_i, b_j are integers $\gtrless 0$, and if there is an infinity of odd integers n, > 0, for which*

$$\sum_{i=1}^{r} a_i^n = \sum_{j=1}^{s} b_j^n,$$

then $s = r$, and the a_i are a permutation of the b_j.

LEMMA 3. *If a_i, $(i = 1, 2, \cdots, r)$ are integers $\gtreqless 0$, and b_j, $(j = 1, 2, \cdots, s)$ integers $\gtrless 0$; and if for all integral values > 0 of n,*

$$\sum_{i=1}^{r} a_i^{2n} = \sum_{j=1}^{s} b_j^{2n},$$

then, (i): $r \geqq s$, *and precisely* $(r - s)$ *of the* $a_i = 0$. (ii) *If, without loss of generality, the s non-zero a_i are a_1, a_2, $\cdots$, a_s, then by Lemma 2, the a_1^2, a_2^2, $\cdots$, a_s^2 are a permutation of the b_1^2, b_2^2, $\cdots$, b_s^2; and hence, by Lemma 1:*

$$\text{(iii)} \quad \sum_{i=1}^{s} f(a_i|) = \sum_{j=1}^{s} f(b_j|); \qquad \sum_{i=1}^{r} f(a_i|) = (r-s) f(0|) + \sum_{j=1}^{s} f(b_j|).$$

The first part is an immediate consequence of Lemma 2.

LEMMA 4. *If a_{kj}, α_{lj}, m_k, μ_k, $(k = 1, \cdots, r;\ j = 1, \cdots, s;\ l = 1, \cdots, t)$, are integers $\gtrless 0$, and if for all integral values > 0 of n_j,*

$$\sum_{k=1}^{r} m_k\, a_{k1}^{2n_1} a_{k2}^{2n_2} \cdots a_{ks}^{2n_s} = \sum_{k=1}^{t} \mu_k\, \alpha_{k1}^{2n_1} \alpha_{k2}^{2n_2} \cdots \alpha_{ks}^{2n_s}, \tag{24}$$

then

$$\sum_{k=1}^{r} m_k f(a_{k1}, a_{k2}, \cdots, a_{ks}|) = \sum_{k=1}^{t} \mu_k f(\alpha_{k1}, \alpha_{k2}, \cdots, \alpha_{ks}|). \tag{25}$$

---

* Lemma 2 was proposed as a problem by the writer in the American Mathematical Monthly; and a proof given ibid., vol. 24 (1917), p. 288, by Professor E. Swift. An independent proof of Lemma 3 is readily deduced from Newton's formulas in the theory of equations, on considering the a_i, b_j as the roots of two equations of degrees r, s respectively, and then showing these equations identical by the given conditions. Hence, (i), (ii) of Lemma 3 being sufficient for the proof of all following Lemmas, it follows that the paraphrase method depends only upon finite processes. The lemmas may be generalized by lightening the conditions; but as such generalizations have no application in the sequel they have been omitted. Professor C. F. Gummer has given (in a paper which has not yet appeared) some interesting developments of Lemma 1, based upon the extension of Descartes' rule of signs to transcendental equations. In particular he has shown that $\Sigma_{i=1}^{i=r} a_i^n = \Sigma_{j=1}^{j=s} b_j^n$ $(r \geqq s)$ for r distinct values of n are necessary and sufficient conditions for the identity of the a_i with the b_j, when n is odd.

Without loss of generality we may assume $m_k, \mu_k > 0$, the other case being immediately reducible by transposition to this. By repeating the terms a proper number of times the coefficients m_k, μ_k may be taken as unity. Now putting $n_i = n\nu_i$, $(i = 1, 2, \cdots, s)$, where ν_i is an arbitrary integer > 0, we infer from (24) by Lemma 3, a set of identities of the form

$$a_{i1}^{2\nu_1} a_{i2}^{2\nu_2} \cdots a_{is}^{2\nu_s} = \alpha_{j1}^{2\nu_1} \alpha_{j2}^{2\nu_2} \cdots \alpha_{js}^{2\nu_s} \quad (1 \leqq i \leqq r;\ 1 \leqq j \leqq t), \tag{25a}$$

valid for all integral values > 0 of the $\nu_1, \nu_2, \cdots, \nu_s$. Replacing in (25a) any one of the exponents by its double, say ν_r by $2\nu_r$, we get an identity, which, with (25a), gives, provided the a's and α's are not zero, for all $\nu_r > 0$, $a_{ir}^{2\nu_r} = \alpha_{jr}^{2\nu_r}$; and hence $a_{ir}^2 = \alpha_{jr}^2$. Hence (25a) gives $(|a_i|)_s = (|\alpha_j|)_s$, and the lemma follows at once by § 21. Obviously the condition $m_k, \mu_k \geqq 0$ may be replaced by $m_k, \mu_k \gtreqless 0$; a remark of importance presently in passing to L-functions of parity $p(0|b_1, b_2, \cdots, b_s)$. We point out expressly that the replacing of the condition $a_{kj}, \alpha_{lj} \geqq 0$ by $a_{kj}, \alpha_{lj} \gtreqless 0$, would invalidate the proof. There is a fundamental distinction between paraphrases involving zero matrices and those which do not. In passing from circular to L-functions, this amounts to distinguishing the paraphrase of homogeneous polynomials in sines and cosines from the paraphrase of the non-homogeneous. We take the former case first.

23. For the a, α, m, μ as in Lemma 4:

LEMMA 5. *If for all values of $x_1, x_2, \cdots, x_s$,*

$$\sum_{k=1}^{r} m_k \cos a_{k1} x_1 \cos a_{k2} x_2 \cdots \cos a_{ks} x_s = \sum_{k=1}^{t} \mu_k \cos \alpha_{k1} x_1 \cos \alpha_{k2} x_2 \cdots \cos \alpha_{ks} x_s, \tag{26}$$

then (25) *holds.*

For, equating coefficients of $x_1^{2n_1} x_2^{2n_2} \cdots x_s^{2n_s}$ in (26) we get (24).

LEMMA 6. *The notation being as in Lemma 5, and for all values of $x_1, x_2, \cdots, x_s$,*

$$\sum_{k=1}^{r} m_k \sin a_{k1} x_1 \sin a_{k2} x_2 \cdots \sin a_{ks} x_s = \sum_{k=1}^{t} \mu_k \sin \alpha_{k1} x_1 \sin \alpha_{k2} x_2 \cdots \sin \alpha_{ks} x_s, \tag{27}$$

then

$$\sum_{k=1}^{r} m_k\, g(|a_{k1}, a_{k2}, \cdots, a_{ks}) = \sum_{k=1}^{t} \mu_k\, g(|\alpha_{k1}, \alpha_{k2}, \cdots, \alpha_{ks}).$$

For, from the definitions in § 1 we may write

$$g(|z_1, z_2, \cdots, z_s) \equiv z_1 z_2 \cdots z_s f(z_1, z_2, \cdots, z_s|).$$

Operating on both sides of (27) with $\partial^s/\partial x_1\,\partial x_2 \cdots \partial x_s$ reduces (27) to (26) with $m_k\,a_{k1} \cdots a_{ks}$, $\mu_k\,\alpha_{k1} \cdots \alpha_{ks}$ in place of m_k, μ_k respectively; and by Lemma 5 we deduce (25) with m_k, μ_k similarly changed:

$$\sum_{k=1}^{r} m_k\,a_{k1}\,a_{k2} \cdots a_{ks} f(a_{k1}, a_{k2}, \cdots, a_{ks}|)$$
$$= \sum_{k=1}^{t} \mu_k\,\alpha_{k1}\,\alpha_{k2} \cdots \alpha_{ks} f(\alpha_{k1}, \alpha_{k2}, \cdots, \alpha_{ks}|).$$

On replacing in this $z_1 z_2 \cdots z_s f(z_1, z_2, \cdots, z_s|)$ by $g(|z_1, z_2, \cdots, z_s)$, the Lemma follows. By an obvious change in notation, Lemma 6 may be restated in the more convenient form:

The a_{ki} denoting integers $\gtrless 0$, and the m_k integers $\gtreqless 0$, the identity in the x_i,

$$\sum_{k=1}^{r} m_k \sin a_{k1} x_1 \sin a_{k2} x_2 \cdots \sin a_{ks} x_s = 0$$

implies

$$\sum_{k=1}^{r} m_k\, g(|a_{k1}, a_{k2}, \cdots, a_{ks}) = 0.$$

Clearly, the preceding Lemmas may be similarly restated. In the same way, the proof of Lemma 7 can be based on Lemma 5 by operating on the identity in Lemma 7 by $\partial^s/\partial y_1\,\partial y_2 \cdots \partial y_s$:

LEMMA 7. *The a_{ki}, b_{kj} denoting integers $\gtrless 0$, and the m_k integers $\gtreqless 0$, the identity in the x_i, y_j,*

$$\sum_{k=1}^{n} m_k \left(\prod_{i=1}^{r} \cos a_{ki}\, x_i \cdot \prod_{j=1}^{s} \sin b_{kj}\, y_j \right) = 0,$$

implies

$$\sum_{k=1}^{n} m_k f(a_{k1}, a_{k2}, \cdots, a_{kr} | b_{k1}, b_{k2}, \cdots, b_{ks}) = 0.$$

A little reflection will show that the method of proof used in Lemmas 6, 7 is applicable when and only when the a_{ki}, b_{kj} are rational.* As there is no essential gain in generality by considering rational rather than integral variables in the paraphrases, we ignore the former.

24. The terms in any one of the trigonometric identities paraphrased in § 23 are all of the same parity. Thus, in Lemma 7, the parity of each sine-cosine term is $p(1^r|1^s)$. Passing to an important generalization we now consider the paraphrase of homogeneous sine-cosine identities whose terms are of several parities. In the following the notation is based upon that of § 6, for the proofs of the processes there stated are intimately connected with Lemma 8, next considered.

* Cf. § 35, footnote.

LEMMA 8. *Let the set of* ω *independent variables,* $z_1, z_2, \cdots, z_\omega$, *be separated in* N *ways into two sets,* X_n, Y_n:

$$X_n \equiv x_{1n}, x_{2n}, \cdots, x_{r_n n}; \qquad Y_n \equiv y_{1n}, y_{2n}, \cdots, y_{s_n n} \quad (n = 1, 2, \cdots, N),$$

so that $r_n + s_n = \omega$, *and the* x_{in}, y_{jn} *are a permutation of the* z_k. *Let all the* X_n, *and consequently all the* Y_n, *be distinct among themselves, two sets being identical only when all the variables in either are also in the other. Write*

$$\phi_m(n) \equiv \prod_{i=1}^{r_n} \cos \alpha_{min}\, x_{in} \cdot \prod_{j=1}^{s_n} \sin \beta_{mjn}\, y_{jn}; \tag{28}$$

$$\psi_m(n) \equiv f_n(\alpha_{m1n}, \alpha_{m2n}, \cdots, \alpha_{mr_n n} \mid \beta_{m1n}, \beta_{m2n}, \cdots, \beta_{ms_n n}); \tag{29}$$

$$\Phi(n) \equiv \sum_{m=1}^{t_n} c_{mn}\, \phi_m(n); \qquad \Psi(n) \equiv \sum_{m=1}^{t_n} c_{mn}\, \psi_m(n). \tag{30}$$

Then, the $\alpha_{min}, \beta_{mjn}$ *denoting integers* $\gtrless 0$, *and the* c_{mn} *integers* $\gtreqless 0$, *the identity in the* z_k,

$$\sum_{n=1}^{N} \Phi(n) = 0 \tag{30a}$$

implies

$$\sum_{n=1}^{N} \Psi(n) = 0, \tag{31}$$

and it will be shown that each term of this sum is zero, viz.,

$$\Psi(n) = 0 \qquad (n = 1, 2, \cdots, N). \tag{31a}$$

For, the X_n being distinct, after operation on (30a) with

$$\partial^{s_n} / \partial y_{1n}\, \partial y_{2n} \cdots \partial y_{s_n n},$$

every $\Phi'(k)$, $k \neq n$ will involve at least one sine factor in each of its terms, $c_{mk}\, \phi_m'(k)$; while each term, $c_{mn}\, \phi_m'(n)$, of $\Phi'(n)$ will be c_{mn} times a product of ω cosines. Hence in the differentiated (30a) only the terms arising from $\Phi(n)$ contribute to the coefficient of $z_1^{2p_1} z_2^{2p_2} \cdots z_\omega^{2p_\omega}$, $p_i > 0, (i = 1, 2, \cdots, \omega)$; and precisely as in the proofs of Lemmas 5, 6, 7, we conclude that $\Psi(n) = 0$, $(n = 1, 2, \cdots, N)$, and hence

$$\sum_{n=1}^{N} \Psi(n) = 0.$$

It is essential for our present purpose to note that (31a) is a system of identities for N general L-functions. That is, $f_1, f_2, \cdots, f_N$ in (31a) may denote the same or different L-functions, which, except that their parities are respectively identical with those of the $\phi_m(n)$, $(n = 1, 2, \cdots, N)$, are wholly arbitrary as defined in § 1.

25. Before proving the general result it will be well for clearness to give the proof in detail for a very simple case, unencumbered by the notation. The reasoning in the general case is of exactly the same kind. We shall now show that

$$\sum c_i \cos(a_i x + b_i y) = 0 \tag{32}$$

for all values of x, y implies

$$\sum c_i f((a_i, b_i)|) = 0, \tag{32a}$$

the c_i denoting integers $\gtreqless 0$, and the a_i, b_i integers $\gtrless 0$. The significance of the parenthesis (a_i, b_i) will be evident on referring to § 1 and the examples of the bar notation there given.

(i) From (32):

$$\sum c_i [\cos a_i x \cos b_i y - \sin a_i x \sin b_i y] = 0;$$

whence, by Lemma 8:

$$\sum c_i f_1(a_i, b_i|) = 0; \qquad \sum c_i f_2(|a_i, b_i) = 0, \tag{33}$$

in which f_1, f_2 are arbitrary, of the indicated parities $p(1^2|0)$, $p(0|1^2)$.

(ii) Now, it is shown* in the proof of the theorem stated in § 3 that the parities $p(1^\alpha|1^\beta)$ of the L-functions appropriate for the stated linear expression of the (general) $f(A_{m1}, A_{m2}, \cdots, A_{mr}|B_{m1}, B_{m2}, \cdots, B_{ms})$ in § 6, whose parity is $p(a_1, a_2, \cdots, a_r|b_1, b_2, \cdots, b_s)$, are precisely those of the several sine-cosine terms in the addition-theorem development and subsequent distribution of products in

$$\prod_{i=1}^{r} \cos\left(\sum_{n=1}^{a_i} \alpha_{min}\, x_{in}\right) \cdot \prod_{j=1}^{s} \sin\left(\sum_{n=1}^{b_j} \beta_{mjn}\, y_{jn}\right). \tag{34}$$

It is shown, moreover, that the appropriate L-functions are of the form (29), corresponding to the individual terms of (34), the latter being of the form (28).

(iii) In the present case we have, therefore, that $f((a_i, b_i)|)$ is a linear function of suitably chosen $f_1(a_i, b_i|)$, $f_2(|a_i, b_i)$, say

$$f((a_i, b_i)|) = k_1 f_1'(a_i, b_i|) + k_2 f_2'(|a_i, b_i).\dagger \tag{35}$$

*Bulletin of the American Mathematical Society, vol. 25 (1918–19), p. 313.

† The actual forms of f_1', f_2' are given by:

$$2f_1'(a_i, b_i|) = f((a_i, b_i)|) + f((a_i, -b_i)|),$$
$$2f_2'(|a_i, b_i) = f((a_i, b_i)|) - f((a_i, -b_i)|); \qquad k_1 = k_2 = 1;$$

but these are not essential to the proof. The same applies to the general case: it is not necessary to have the linear expressions; it is sufficient to know that they exist.

Multiplying (35) throughout by c_i, and summing:

$$(35a) \qquad \sum c_i f((a_i, b_i)|) = k_1 \sum c_i f_1'(a_i, b_i|) + k_2 \sum f_2'(|a_i, b_i).$$

But (33) holds for f_1, f_2 arbitrary of the indicated parities. Hence the right side of (35a) vanishes, and this establishes (32a).

26. Turning to § 6, we may write (8) in the form (30a) by the process outlined in § 25 (ii). In this case $c_{mn} \equiv c_m$; and it is easy to see that $N = 2^{\omega-\delta}$, where ω, δ are as in § 2, (5). By Lemma 8 we get from (30a), corresponding to (31a):

$$(36) \qquad \Psi(n) = 0 \qquad (n = 1, 2, 3, \cdots, 2^{\omega-\delta}).$$

Choosing for the f_n in (36) the L-functions f_n' appropriate for the linear expressions (§ 3) of $f(A_{m1}, A_{m2}, \cdots, A_{mr}|B_{m1}, B_{m2}, \cdots, B_{ms})$, and multiplying as in § 25 (iii) the successive identities of (36) by the appropriate constants, k_n, of the linear expression, we get on adding the results as in the special (35a), the identity (8a) of § 6.

27. The proofs for (9a), (10a) in the homogeneous case are precisely similar to that for (8a), and need not be written out. Again we emphasize that (8a), (9a), (10a) have so far been proved only for the case in which the α_{min}, β_{mjn} are non-zero integers. We next (cf. § 22) consider in less detail the paraphrase process for non-homogeneous sine-cosine polynomials. We shall give only so much of it as suffices for the paraphrases of identities first arising in the elliptic and theta functions; this includes all of the Liouville paraphrases and many more of kinds distinct from his. The most general non-homogeneous case may be similarly treated, but the notation becomes considerably more complicated, and it is best, by using the linear transformations outlined in § 35, to refer back to the homogeneous case. By the method of sets, much information not otherwise evident, is revealed concerning the ultimate nature of the paraphrases.

LEMMA 9. *The identity in* x_1, x_2:

$$(37) \qquad \begin{aligned} &\sum_{i=1}^{r} (\cos a_{i1} x_1 \cos a_{i2} x_2 - \cos a_{i3} x_1 \cos a_{i4} x_2) \\ &\qquad = \sum_{i=1}^{s} (\cos a_{i5} x_2 - \cos a_{i6} x_1) \end{aligned}$$

where the a's are integers, and $a_{i2}, a_{i3}, a_{i5}, a_{i6} \gtrless 0$, *implies*

$$\sum_{i=1}^{r} [f(a_{i1}, a_{i2}|) - f(a_{i3}, a_{i4}|)] = \sum_{i=1}^{s} [f(0, a_{i5}|) - f(a_{i6}, 0|)].$$

For, equating coefficients of $x_1^{2n}, x_2^{2n}, x_1^{2n_1}, x_2^{2n_1}, (n, n_1, n_2 > 0)$ in (37),

we get:

$$\sum_{i=1}^{r} a_{i1}^{2n} + \sum_{i=1}^{s} a_{i6}^{2n} = \sum_{i=1}^{r} a_{i3}^{2n}, \tag{38}$$

$$\sum_{i=1}^{r} a_{i4}^{2n} + \sum_{i=1}^{s} a_{i5}^{2n} = \sum_{i=1}^{r} a_{i2}^{2n}, \tag{39}$$

$$\sum_{i=1}^{r} a_{i1}^{2n_1} a_{i2}^{2n_2} = \sum_{i=1}^{r} a_{i3}^{2n_1} a_{i4}^{2n_2} \tag{40}$$

for all integral $n, n_1, n_2 > 0$. Since* the a's, except perhaps some of the a_{i1}, a_{i4}, are not zero, we find from (38), (39) and Lemma 3 that precisely s each of the a_{i1}, a_{i4} are zero. Moreover as in Lemma 4 we find that in (40) *the pairs* $(|a_{i1}|, |a_{i2}|)$ *for which* $a_{i1} \neq 0$ *are merely a permutation of the pairs* $(|a_{i3}|, |a_{i4}|)$ *for which* $a_{i4} \neq 0$. Hence if $a_{i1} = 0$, then $a_{i4} = 0$. Suppose this true for $i = 1, \cdots, s$. Then

$$\sum_{i=1}^{r} [f(a_{i1}, a_{i2}|) - f(a_{i3}, a_{i4}|)] = \sum_{i=1}^{s} [f(0, a_{i2}|) - f(a_{i3}, 0|)].$$

But again from (38), (39) and Lemma 3 we see that the first s of the $|a_{i2}|$ are the $|a_{i5}|$, and the first s of the $|a_{i3}|$ are the $|a_{i6}|$, which proves the paraphrase. From this there is obviously the corollary:

LEMMA 10. *With the notation of Lemma 9, and b_i, c_i integers $\geqq 0$,*

$$\sum_{i=1}^{r} b_i(\cos a_{i1}x_1 \cos a_{i2}x_2 - \cos a_{i3}x_1 \cos a_{i4}x_2) = \sum_{i=1}^{s} c_i(\cos a_{i5}x_2 - \cos a_{i6}x_1) \tag{45}$$

for all values of x_1, x_2 implies

$$\sum_{i=1}^{r} b_i[f(a_{i1}, a_{i2}|) - f(a_{i3}, a_{i4}|)] = \sum_{i=1}^{s} c_i[f(0, a_{i5}|) - f(a_{i6}, 0|)]. \tag{46}$$

It is not difficult to prove this also in the case $b_i, c_i \gtreqless 0$; but this is not an immediate consequence of Lemma 9.

28. By a process of frequent use we get from Lemma 10 an important special case as a corollary. Obviously (46) is true for all integers for which (45) is true. But (45) is true for the integers $a_{i2} = a_{i4} = a_{i5} = 0$ (the other integers being the same), since this is the form which (45) takes when $x_2 = 0$. Hence

LEMMA 11. *The identity in x_1:*

$$\sum_{i=1}^{r} b_i(\cos a_{i1}x_1 - \cos a_{i2}x_1) = \sum_{i=1}^{s} c_i(1 - \cos a_{i3}x_1)$$

* An alternative proof by the method of sets is somewhat longer and has been omitted, but is not without interest. It may easily be reconstructed from (38)–(40), Lemma 3, and § 20 (21), (22). (To save renumbering formulas, (45) follows (40).)

implies

$$\sum_{i=1}^{r} b_i[f(a_{i1}|) - f(a_{i2}|)] = \sum_{i=1}^{s} c_i[f(0|) - f(a_{i3}|)].$$

This* may be proved independently by Lemma 9; or it follows almost at once from Lemma 3. It is of interest as covering the first paraphrase stated by Liouville, which follows from Jacobi's series for $\mathrm{sn}^2 u$ from the identity $\mathrm{sn}\, u \times \mathrm{sn}\, u = \mathrm{sn}^2 u$, on substituting for $\mathrm{sn}\, u$, $\mathrm{sn}^2 u$ their Fourier developments. The generalizations to functions of two variables in Liouville's first five memoirs follow from Lemmas 9, 10 applied to the appropriate series, which also were given by Jacobi, but not in the *Fundamenta Nova*. The formulas of Liouville's sixth memoir are paraphrases of

$$\mathrm{sn}^3 u = \mathrm{sn}^2 u \times \mathrm{sn}\, u = \mathrm{sn}\, u \times \mathrm{sn}\, u \times \mathrm{sn}\, u.$$

29. By differentiation as in §§ 23, 24 we may make the cases of non-homogeneous paraphrases for functions of parity $p(0|2)$, $p(0|a)$, $\cdots$ depend upon those for functions of parity $p(2|0)$, $p(a|0)$, $\cdots$. We shall consider it unnecessary to prove formally the legitimacy of paraphrasing non-homogeneous identities differing but slightly from those considered in §§ 27, 28; and for the present we may omit the paraphrase of identities involving tangents, cotangents, secants and cosecants, these depending upon sixteen simple identities which will be given with the elliptic series, and in no respect introducing considerations different in principle from the paraphrase of sine-cosine identities. We remark, however, that they are the source of all such paraphrases as those of Liouville which involve sums of L-functions whose arguments are in arithmetical or geometrical progression, such, for instance, as (14), (19), (19a).

III. Elementary transformations

30. Examining Liouville's theorems we note his frequent use of such transformations as $f_1(z|) = (-1)^{(z+1)/2} f_2(|z)$, where z is an odd integer, and f_1, f_2 arbitrary of the parities indicated. These are immediate translations of the effects of replacing the x-variables in the elliptic or theta identities from

* G. Humbert, (Paris Comptes Rendus, vol. 150; 21 Fév. 1910, p. 433) uses what is essentially a special case of Lemma 11, and refers for proof to a theorem of Borel: "*There exists an entire function of x, taking for integral values of the variable the same values as any given function.*" Liouville functions being not necessarily entire, Borel's theorem cannot be used to prove Lemma 11; and in any event it is preferable here to use some method, such as the above, which is applicable to functions of any number of variables. On the other hand, some writers (cf. Bachmann, l. c., p. 295), regard the paraphrase to functions of a single variable as self-evident. Our lemmas are, no doubt, obvious; but in view of the indicated difference of opinion as to what is or is not obvious in this regard, it seemed best to offer proofs for all cases.

which the paraphrases are derived by $x + \pi/2$, or in Weierstrass' notation, by $x + 1/2$. In the notation of § 7, all such transformations follow from that next given, which may be verified by inspection. The functions in any pair are of the same parity, and the sign of transformation, $\sim$, indicates that in any paraphrase either function separated by the sign $\sim$ may be replaced throughout by the other, provided, of course, that the evenness or oddness of the integral arguments of the functions in the paraphrase is constant throughout. Thus, $\sum_i f(|m_i, 2n_i) = 0$ may be replaced by $\sum_i (-1)^{(m_i-1)/2} f(m_i | 2n_i) = 0$;

$$\sum_i -(1)^{n_i} f(|m_i, 2n_i) = 0; \qquad \sum_i (-1)^{[n_i+(m_i-1/2)]} f(m_i|2n_i) = 0.$$

It is readily seen that if

$$\xi \equiv (m_1, m_2, \cdots, m_r, m_1', m_2', \cdots, m_s', 2n_1, 2n_2, \cdots, 2n_k, 2n_1', 2n_2', \cdots, 2n_i');$$

$$M = m_1 + m_2 + \cdots + m_r; \qquad N = n_1 + n_2 + \cdots + n_k:$$

then

$$f(|\xi) \sim (-1)^{(M-1)/2} f(\xi|), \quad f(\xi|) \sim (-1)^{(M+1)/2} f(|\xi), \quad \text{if } M \equiv 1 \bmod 2;$$

$$f(|\xi) \sim (-1)^{M/2} f(|\xi), \qquad f(\xi|) \sim (-1)^{M/2} f(\xi|), \qquad \text{if } M \equiv 0 \bmod 2;$$

$$f(\xi|) \sim (-1)^N f(\xi|), \qquad f(|\xi) \sim (-1)^N f(|\xi).$$

31. We may regard $f(\xi_1, \xi_2, \cdots, \xi_r | \eta_1, \eta_2, \cdots, \eta_s)$, the ξ, η being the matrices of § 1, as an L-function of ξ_1 alone, or of η_1 alone. Hence in the following sections we need consider only the behavior of functions of parity $p(a|0)$, $p(0|b)$, and need examine paraphrases for functions of those parities alone. By repeated application of the theorems below for functions of parity $p(a|0)$ and $p(0|b)$, the results for functions of parity

$$p(a_1, a_2, \cdots, a_r | b_1, b_2, \cdots, b_s)$$

may be written out if desired, cf. § 6.

32. Let $\xi_i \equiv (\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{ia})$, $\eta_i \equiv (\beta_{i1}, \beta_{i2}, \cdots, \beta_{ib})$. Then the matrices $(\xi_i; \eta_i)$, $(\xi_i; -\eta_i)$, where

$$(\xi_i; \eta_i) \equiv (\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{ia}, \beta_{i1}, \beta_{i2}, \cdots, \beta_{ib}),$$

$$(\xi_i; -\eta_i) \equiv (\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{ia}, -\beta_{i1}, -\beta_{i2}, \cdots, -\beta_{ib}),$$

are termed the conjoints of ξ_i, η_i, of ξ_i, $-\eta_i$ respectively.

Consider now a paraphrase (over a given separation):

$$\sum_i a_i f((\xi_i; \eta_i)|) = 0. \tag{47}$$

Choose for $f((\xi_i; \eta_i)|)$ the implicitly restricted L-function:

$$\cos\left(\sum_{r=1}^{a} \alpha_{ir} x_r + \sum_{s=1}^{b} \beta_{is} y_s\right) \tag{48}$$

in which the x, y are parameters. Substituting (48) in (47), and applying Lemma 8 (§ 24), we infer by § 3, as in § 25, for f_1, f_2 arbitrary of the indicated parities $p(a, b|)$, $p(|b, a)$:

$$(49) \qquad \sum a_i f_1(\xi_i, \eta_i|) = 0; \qquad \sum a_i f_2(|\xi_i, \eta_i) = 0,$$

as consequences of (47). Similarly, as consequences of

$$(50) \qquad \sum_i a_i f(|(\xi_i; \eta_i)) = 0,$$

we find in precisely the same way:

$$(51) \qquad \sum_i a_i f_1(\xi_i|\eta_i) = 0; \qquad \sum_i a_i f_2(\eta_i|\xi_i) = 0.$$

33. The results of § 32 are paraphrases, ultimately, of the addition theorems for the sine and cosine. So also are the following obvious identities, which are frequently useful, cf. § 3.

$$(52) \qquad \begin{aligned} f((\xi_i; \eta_i)|) &= f_1(\xi_i, \eta_i|) - f_2(|\xi_i, \eta_i), \\ f(|(\xi_i; \eta_i)) &= f_3(\eta_i|\xi_i) + f_4(\xi_i|\eta_i); \end{aligned}$$

where

$$(53) \qquad \begin{aligned} 2f_1(\xi_i, \eta_i|) &\equiv f((\xi_i; -\eta_i)|) + f((\xi_i; \eta_i)|), \\ 2f_2(|\xi_i, \eta_i) &\equiv f((\xi_i; -\eta_i)|) - f((\xi_i; \eta_i)|), \\ 2f_3(\eta_i|\xi_i) &\equiv f(|(\xi_i; \eta_i)) + f(|(\xi_i; -\eta_i)), \\ 2f_4(\xi_i|\eta_i) &\equiv f(|(\xi_i; \eta_i)) - f(|(\xi_i; -\eta_i)). \end{aligned}$$

That f_1, f_2, f_3, f_4 have the parities implied by their bar notations may be verified at once from the definitions of § 1.

34. Obviously $f(\xi_1, \xi_1, \cdots, \xi_1, \xi_2, \xi_3, \cdots, \xi_r|)$ is no more general than $f(\xi_1, \xi_2, \cdots, \xi_r|)$. Similarly any L-function may be formally reduced by omitting from its symbol redundant matrices. This obvious remark will appear in the sequel as the source of some of Liouville's most difficult paraphrases (from the standpoint of proof of Dirichlet's method). We next consider the complement of this process of reduction. It leads from paraphrases for functions of order ω to paraphrases in which the order of the function exceeds ω, again a process which seems to have been employed by Liouville to transform his simpler results.

35. To keep the writing simple, we may at this stage confine our attention to functions of order 2 integrated over separations of degree 3, deferring the general case, which is treated in the same way, and the theory of classes of paraphrases until we shall have in the next paper a considerable body of theorems for particular functions and separations by which to illustrate the processes involved. For simplicity, since we are to consider functions of

order 2, choose for the F of § 8 (13), $F \equiv f((z_1, z_2)|)$. The partition is to be of degree 3; hence in the notation of § 8, where now (cf. footnote) the l's are integral,

$$\Lambda_i = l_{i1}\lambda_1 + l_{i2}\lambda_2 + l_{i3}\lambda_3 \qquad (i = 1, 2);$$

and for the paraphrase (13), we have in the present case:

$$(54) \quad \sum a_i f((l_{11}\lambda_1 + l_{12}\lambda_2 + l_{13}\lambda_3, l_{21}\lambda_1 + l_{22}\lambda_2 + l_{23}\lambda_3)|) = 0,$$

the $\sum$ extending over all $\lambda_1, \lambda_2, \lambda_3$ defined by the separation. Write

$$(55) \quad A \equiv \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r; \qquad B \equiv \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r,$$

the x's denoting parameters, and the α, β constant integers.* As in § 32, replace (54) by its special case:

$$(56) \quad \sum a_i \cos\{(l_{11}\lambda_1 + l_{12}\lambda_2 + l_{13}\lambda_3)\zeta_1 + (l_{21}\lambda_1 + l_{22}\lambda_2 + l_{23}\lambda_3)\zeta_2\} = 0,$$

an identity in ζ_1, ζ_2. Substitute for the parameters ζ_1, ζ_2 in (56), A, B, respectively; then there is the identity in the x's:

$$(57) \quad \sum a_i \cos(L_1 x_1 + L_2 x_2 + \cdots + L_r x_r) = 0,$$

where

$$(58) \quad L_i \equiv (\alpha_i l_{11} + \beta_i l_{21})\lambda_1 + (\alpha_i l_{12} + \beta_i l_{22})\lambda_2 + (\alpha_i l_{13} + \beta_i l_{23})\lambda_3;$$

and (57) paraphrases into:

$$(59) \quad \sum a_i f((L_1, L_2, \cdots, L_r)|) = 0.$$

By suitably choosing the constants α_i, β_i, the L_i may be taken equal to linear functions of the $\lambda_1, \lambda_2, \lambda_3$, to a certain extent predetermined; viz., if $L_i = l_i\lambda_1 + m_i\lambda_2 + n_i\lambda_3$, the values of any two of the l_i, m_i, n_i fix the value of the third. Applying § 32 to (59), we deduce from it paraphrases for functions of parity $p(r_1|r_2)$, where $r_1 + r_2 = r$.

If we had chosen $F \equiv f(z_1|z_2)$, we should have had in place of (56):

$$(56a) \quad \sum a_i \cos(l_{11}\lambda_1 + l_{12}\lambda_2 + l_{13}\lambda_3)\zeta_1 \cdot \sin(l_{21}\lambda_1 + l_{22}\lambda_2 + l_{23}\lambda_3)\zeta_2 = 0;$$

whence

$$\sum a_i [\sin\{(l_{21}\zeta_2 + l_{11}\zeta_1)\lambda_1 + (l_{22}\zeta_2 + l_{12}\zeta_1)\lambda_2 + (l_{23}\zeta_2 + l_{13}\zeta_1)\lambda_3\}$$
$$+ \sin\{(l_{21}\zeta_2 - l_{11}\zeta_1)\lambda_1 + (l_{22}\zeta_2 - l_{12}\zeta_1)\lambda_2 + (l_{23}\zeta_2 - l_{13}\zeta_1)\lambda_3\}] = 0,$$

* There is no difficulty in extending this to the case of α, β numerical constants from any field. A like remark applies to the lemmas of §§ 22–28. In particular, if the α, β in F, G, H of § 6 denote rational numbers, it is obviously possible on replacing the x, y variables in (8), (9), (10) by suitable integral multiples of themselves, to reduce (8), (9), (10) to forms in which the α, β are replaced by integers, and the paraphrases of these forms may be taken by definition as the equivalents of the paraphrases of the first forms in which the α, β were rational. The cases of transcendental α, β or α, β belonging to other fields are ignored because non-trivial identities (8), (9), (10) involving such numbers do not yet (apparently) exist. Cf. § 23.

and the work henceforth is of the same kind as above. It is an interesting exercise on this section and the next to show that (19) is transformed into (19*a*) by the substitution

$$x \sim \tfrac{1}{2}x + \tfrac{1}{2}y + z, \qquad y \sim \tfrac{1}{2}x - \tfrac{1}{2}y + w.$$

36. Without considering explicitly restricted *L*-functions in detail at this point, we may illustrate their origin by a simple example. Again the general case is of the same nature, and the work for it similar to that for the special example. The *L*-function

$$f(x, y|) - f(y, x|), \qquad \equiv \phi(x, y|) \tag{60}$$

obviously satisfies $\phi(x, y|) = -\phi(y, x|)$. Conversely, if it be required to determine the form of the most general *L*-function, $\psi(x, y)$, of parity $p(1^2|0)$, which changes sign with interchange of the variables, we have, expressing the parity conditions, $\psi(x, y) \equiv F(x, y|)$; and, by the given condition, $\psi(y, x) \equiv F(y, x|) = -\psi(x, y)$. Whence

$$2\psi(x, y) = F(x, y|) - F(y, x|);$$

F being unrestricted of parity $p(1^2|0)$. An arbitrary constant factor may clearly be absorbed in an *L*-function without changing its parity or diminishing its generality; hence, we may take $F(x, y|) = 2f(x, y|)$, f arbitrary of parity $p(1^2|0)$; and $\psi(x, y) \equiv \phi(x, y|)$.

The forms of restricted functions which it is profitable to investigate are suggested by the elliptic and theta identities. One of the chief uses of restricted *L*-functions is to sum up in compendious form paraphrases for unrestricted *L*-functions. Thus, the paraphrase* $\sum a_i[f(x_i, y_i|) - f(y_i, x_i|)]$, may be replaced by $\sum a_i \phi(x_i, y_i|) = 0$ where $\phi(x, y|) = -\phi(y, x|)$. Restricted paraphrases may be found directly from the elliptic or theta identities by permuting the variables, multiplying the results by ± 1 and adding and simplifying; or in many other ways that suggest themselves as we proceed. Illustrative of the first method, it may be verified without difficulty that the multitude of paraphrases to which Weierstrass' "equation of three terms" gives rise, are all equivalent to the following, or to special cases of it:

$$\begin{gathered} 4n = m_1 + m_2 + m_3 + m_4; \qquad m_i = d_i\,\delta_i \qquad (i = 1, 2, 3, 4): \\ \sum \phi((d_1 - d_2, \delta_1 + \delta_2, d_3 - d_4, \delta_3 + \delta_4)|) = 0, \end{gathered} \tag{61}$$

where $\phi((x, y, z, w)|) = \phi((x, y, -z, -w)|)$, and $\phi((x, y, z, w)|)$ changes sign under each of the 12 odd substitutions on x, y, z, w.

37. Liouville (11; p. 301, (σ)) has given one example of a paraphrase in-

* § 15 (19) comes under this case.

volving a wholly arbitrary function of a single variable. By the present methods such paraphrases may be found for arbitrary functions of n variables.* For, let $f(x_1, x_2, \cdots, x_n)$ denote an arbitrary function, then:

(62) $2f(x_1, x_2, \cdots, x_n) \equiv f_1((x_1, x_2, \cdots, x_n)|) + f_2(|(x_1, x_2, \cdots, x_n))$,

where f_1, f_2 are given by:

$$(63)\quad \begin{aligned} &f_1((x_1, x_2, \cdots, x_n)|) \\ &\qquad = f(x_1, x_2, \cdots, x_n) + f(-x_1, -x_2, \cdots, -x_n), \\ &f_2(|(x_1, x_2, \cdots, x_n)) \\ &\qquad = f(x_1, x_2, \cdots, x_n) - f(-x_1, -x_2, \cdots, -x_n). \end{aligned}$$

Hence, if by any means we have deduced

$$(64)\quad \begin{aligned} &\sum c_i F((a_{i1}, a_{i2}, \cdots, a_{in})|) = 0, \\ &\sum c_i G(|(a_{i1}, a_{i2}, \cdots, a_{in})) = 0, \end{aligned}$$

in which F is arbitrary of parity $p(n|0)$, G arbitrary of parity $p(0|n)$, we may choose $F \equiv f_1$, $G \equiv f_2$, and by (62) infer

$$(65)\quad \sum c_i f(a_{i1}, a_{i2}, \cdots, a_{in}) = 0.$$

Pairs of paraphrases such as (64) are furnished by the elliptic and theta expansions; hence also paraphrases of the kind (65).

38. Returning for a moment to § 35, we shall illustrate the use of linear transformations in non-homogeneous paraphrases by giving an alternative proof of Lemma 9. The general case admits of similar treatment. Writing $x_1 = l_1 x + m_1 y$, $x_2 = l_2 x + m_2 y$ in (37), we infer, as in § 32 (49), l_1, m_1, l_2, m_2 denoting arbitrary integral constants:

$$\begin{aligned} &\sum_{i=1}^{r} [f(l_1 a_{i1} + l_2 a_{i2}, m_1 a_{i1} + m_2 a_{i2}|) + f(l_1 a_{i1} - l_2 a_{i2}, m_1 a_{i1} - m_2 a_{i2}|) \\ &\quad - f(l_1 a_{i3} + l_2 a_{i4}, m_1 a_{i3} + m_2 a_{i4}|) - f(l_1 a_{i3} - l_2 a_{i4}, m_1 a_{i3} - m_2 a_{i4}|)] \\ &\qquad\qquad = 2 \sum_{i=1}^{s} [f(l_2 a_{i5}, m_2 a_{i5}|) - f(l_1 a_{i6}, m_1 a_{i6}|)]. \end{aligned}$$

Setting in this $l_1 = m_2 = 1$, $l_2 = m_1 = 0$, we find:

$$\sum_{i=1}^{r} [f(a_{i1}, a_{i2}|) - f(a_{i3}, a_{i4}|)] = \sum_{i=1}^{s} f(0, a_{i5}|) - f(a_{i6}, 0|)],$$

as stated in Lemma 9.

* Such paraphrases do not appear to be numerous for the elliptic functions. On the other hand they are of universal occurrence for the theta functions of more than one variable. An account of Kummer's surface from this point of view will be published elsewhere.

We have merely sketched a few of the principal transformation processes, which will be more fully developed when we have written out the elliptic and theta series in a form suitable for paraphrase, to which we pass next, translating as we go the results into paraphrases of the kind described in this paper.

REFERENCES*

* Of this list 2.1, 3.2, 3.3, 3.4, 3.5, 3.6, 9, 10 were supplied by Professor A. J. Kempner from the galleygraphs of chapter eleven of Dickson's 2d vol. of the "*History of the Theory of Numbers.*" With the exception of 2.1, all of these have been inaccessible to me. Quoting from Dickson, Kempner says in regard to 3.6, " N. V. Bougaief proved some of the theorems in Liouville's series of articles by showing that, if $F(x)$ is an even function, an identity $\Sigma_{m=0}^{\infty} A_m \cos mx = \Sigma_{n=0}^{\infty} B_n \cos nx$ implies $\int A_m F(m) = \int B_n F(n)$, and a similar theorem involving sines and an odd function $F_1(n)$." This would appear to be in accordance with Liouville's suggestions, cf. § 13, especially footnote.

1. J. Liouville (1–18): *Sur quelques formules générales qui peuvent être utiles dans la théorie des nombres.* Eighteen memoirs, as follows: 1 to 6 in Journal de mathématiques pures et appliquées (2), vol. 3 (1858); 7 to 11 in vol. 4 (1859); 12 in vol. 5 (1860); 13 to 16 in vol. 9 (1864); 17, 18 in vol. 10 (1865). These will be referred to by citing the number of the memoir and the page.

2. J. Liouville (19): Réponse de M. Liouville; Note de M. Liouville; ibid., vol. 7.

2.1. V. A. Lebesgue, ibid., vol. 7.

3. H. J. S. Smith: *Report on the Theory of Numbers* (1865), (Collected Papers I), Art. 136

3.1. C. M. Piuma, Giornale di Matematiche, vol. 4 (1866), 3 articles.

3.2. E. Fergola, Giornale di Matematiche, vol. 10 (1872).

3.3. G. Torelli, ibid., vol. 16 (1878).

3.4. S. J. Baskakov, Transactions of the Moscow Mathematical Society, vol. 10, I (1882–3).

3.5. T. Pepin, Accademia pontificia dei nuovi Lincei, Atti, vol. 38 (1884–5).

3.6. N. V. Bougaief, Transactions of the Moscow Mathematical Society, vol. 12 (1885).

4. T. Pepin: *Sur quelques formules d'Analyse utiles dans la Théorie des nombres,* Journal de mathématiques pures et appliquées, (4), vol. 4 (1888), pp. 83–131.

5. T. Pepin: *Sur quelques formes quadratiques quaternaires,* ibid., vol. 6 (1890), pp. 1–67. This memoir illustrates one aspect of the usefulness of Liouville's methods.

6. G. B. Mathews: *On a theorem of Liouville's,* Proceedings of the London Mathematical Society, vol. 25 (1893), pp. 85–92.

7. W. Meissner: *Inaugural Dissertation,* Zürich, 1907. Followed by Bachmann in (8).

8. P. Bachmann: *Niedere Zahlentheorie, zweiter Teil, Additive Zahlentheorie,* Leipzig (1910), pp. 365–433.

9. G. Humbert, Bulletin des Sciences Mathématiques (2), vol. 34, I (1910).

10. A. Deltour, Nouvelles Annales de Mathématiques (4), vol. 11 (1911).

UNIVERSITY OF WASHINGTON

THE CONSTRUCTION OF ALGEBRAIC CORRESPONDENCES BETWEEN TWO ALGEBRAIC CURVES*

BY

VIRGIL SNYDER AND F. R. SHARPE

1. **Statement of the problem.** Given two algebraic curves, $C(x_1, x_2, x_3) \equiv C(x) = 0$ of genus p in the plane (x), and $C'(x') = 0$ of genus p' in the plane (x'). Suppose that to a point (y) on C correspond n' points (y') on C', and that to a point (y') on C' correspond n points (y) on C. The two curves C, C' are then said to be in (n, n') correspondence. It is the purpose of this paper to give some methods of constructing curves having such correspondences, and of obtaining the equations which define them.

For certain positions of the point (y), two of the n' images on C' may coincide. Such a point is called a branch-point, and the image point that is counted twice is called a coincidence. If the number of branch-points on C is denoted by η, and on C' by η', then we have by Zeuthen's formula

$$\eta' - \eta = n'(2p - 2) - n(2p' - 2).$$

2. **Intermediary curve.** Let C, C' lie in different planes in ordinary space. Connect each point (y) of C with all the corresponding points (y') on C' by means of straight lines; similarly, connect each point (y') on C' with all its image points on C. In this way a ruled surface R is generated, having C for curve of multiplicity n' and C' of multiplicity n. Let K be an arbitrary plane section of R, and let P be its genus. Through any point of K passes one and in general only one generator g, and this generator meets C in one point. To this point on C correspond n' points on K, namely, the points in which the n' generators through the given point on C meet the curve K. Moreover, all the n' points on K have just this one image on C. The curves C, K are therefore in $(1, n')$ correspondence. Similarly, C', K are in $(1, n)$ correspondence. Hence K has two involutions, one of order n, genus p', the other of order n', genus p.

A branch-point on C gives rise to a branch-point on K, but since to a point on K corresponds only one point on C, there can be no coincidences. By

* Presented to the Society, September 6, 1920.

applying Zeuthen's formula we therefore have

$$2P - 2 = n'(2p - 2) + \eta = n(2p' - 2) + \eta'.$$

We may therefore state the following known

THEOREM: *Associated with every* (n, n') *correspondence between two curves of genera* p, p' *is an intermediary curve of genus* P, *on which exist two involutions, one of order* n', *genus* p, *and another of order* n, *genus* p'.

3. **Series contained in a linear series.** It has been found by Castelnuovo* that the maximum value of η is $2n(n' + p' - 1)$ and of η' is $2n'(n + p - 1)$ and that the maximum values are attained simultaneously. In this case the genus P of K has its maximum value. It was also shown that the necessary and sufficient condition that the (n, n') correspondence between the given curves can be expressed by means of one auxiliary equation is that η or η' attains its maximum value. Let this equation be $\phi(x, x') = 0$.

When (x') is fixed on C', $\phi(x, x') = 0$ defines a curve in the plane (x) which meets C in n points, images of the given point on C'. Similarly, when (x) is fixed on C, the curve $\phi = 0$ meets C' in n' points.

In the older literature no other forms of correspondence were known than this Cayley-Brill theory of correspondence, which is a generalization of the correspondence between two straight lines, as developed by Chasles.† It was pointed out by Hurwitz‡ that not all correspondences can be expressed by means of one equation but that every correspondence can be expressed by means of at most two auxiliary equations

$$(1) \qquad \phi_1(x, x') = 0, \qquad \phi_2(x, x') = 0.$$

No illustrations are given, nor any properties discussed of correspondences requiring two equations for their definition; such correspondences are called *singular* correspondences. The two equations (1) define a multiple correspondence between two planes, having the restricted property that the entire image of C is C' taken multiply, and similarly all the images of points on C' lie on C. Our problem is thus equivalent to that of finding such correspondences.

4. **Intersection of two ruled surfaces.** An example of a curve K having two involutions was given by Amodeo§ and cited by Castelnuovo,|| namely,

* *Sulle serie algebriche di gruppi di punti appartenenti ad una curva algebrica*, Rend. d. R. Accademia dei Lincei, ser. 5, vol. 15(1) (1906), pp. 337–344.

† See Clebsch-Lindemann, *Vorlesungen ueber Geometrie*, vol. 1, p. 437 ff.

‡ *Ueber algebraische Correspondenzen und das verallgemeinerte Correspondenzprincip*, Mathematische Annalen, vol. 28 (1887), pp. 561–593.

§ F. Amodeo, *Contribuzione alla teoria delle serie irrazionale involutorie giacenti sulle varieta algebriche ad una dimensione*, Annali di Matematica, ser. 2, vol. 20 (1892), pp. 229–235.

|| L. c., p. 342.

the curve of intersection of two ruled surfaces in general position. Noether* had proved that the intersection K of R_m, of genus π with $R_{m'}$, of genus π' has the genus P defined by

$$P = (m-1)(m'-1) + m\pi + m'\pi'.$$

Since m, π are precisely the order and genus of one involution on K, and m', π' are the order and genus of the other, it follows from Castelnuovo's theorem that P has its maximum value and therefore that the correspondence can be expressed by one equation. Analytically the equations of a generator of R may be written in the form $\Sigma a_i x_i = 0$, $\Sigma b_i x_i = 0$ where a_i and b_i are rational functions of parameters $\lambda_1, \lambda_2, \lambda_3$ connected by an algebraic relation $f(\lambda_1, \lambda_2, \lambda_3) = 0$ of genus π. Similarly, for R' we have $\Sigma c_i x_i = 0$, $\Sigma d_i x_i = 0$, where c_i, d_i are functions of (λ'), and $f'(\lambda') = 0$.

The condition that the generators intersect is

$$\Delta \equiv (a_1, b_2, c_3, d_4) = 0.$$

This equation expresses the correspondence between

$$f(\lambda_1, \lambda_2, \lambda_3) = 0 \quad \text{and} \quad f'(\lambda_1', \lambda_1', \lambda_3') = 0.$$

5. **Ruled surfaces with common generators.** For certain sets of values of (λ) and of (λ') it may happen that the four planes (a), (b), (c), and (d) have a line in common, instead of simply a point of K. Such a line is a common generator of R and R'. For each common generator it can be proved that the genus of K is reduced by unity, but since each common generator counts for two coincidences, Castelnuovo's condition is still satisfied, and one equation is sufficient to determine the correspondence.

6. **Ruled surfaces with common plane section $\overline{C}$.** Since any plane section of R or of R' is in $(1, 1)$ correspondence with $\overline{C}$, the correspondence is equivalent to a correspondence on $\overline{C}$. A generator of R through a point (y) of $\overline{C}$ meets the residual curve K in points through each of which passes a generator of R', and this generator meets $\overline{C}$ in an image point (y'). We have thus a correspondence of valence 1.

Any case of this kind is an example of the type

$$f(x_1, x_2, x_3) = 0, \tag{2}$$

$$f(x_1', x_2', x_3') = 0, \tag{3}$$

$$A(x_1 x_3' - x_1' x_3) + B(x_2 x_3' - x_2' x_3) = 0. \tag{4}$$

If between (2) and (4) we eliminate x_1, then by means of (3) the factor $x_2 x_3'$

---

* M. Noether, *Zur Theorie des eindeutigen Entsprechens algebraischer Gebilde*, Mathematische Annalen, vol. 8 (1875), pp. 495–533.

$-x_2' x_3$ can be removed; we thus obtain a fourth relation, which completes the definition of the correspondence.

Example: Thus, the two ruled surfaces

$$R \equiv x_2^2 x_3 (x_1 + x_4) = x_1^2 (x_1^2 + 2x_1 x_4 + x_4^2 + x_3^2),$$
$$R' \equiv x_1 x_2^2 x_3 = (x_1 + x_4)^2 (x_1^2 + x_3^2)$$

pass through the curve $f \equiv x_2^2 x_3 - x_2(x_1^2 + x_3^2) = 0$, $x_4 = 0$. The residual curve of intersection of R and R' defines the $(2, 2)$ correspondence on f expressed by

$$x_1'^2 x_2 x_3 - x_1^2 x_2' x_3' = 0, \qquad x_2 x_3 x_1'^2 + x_1 x_3 x_1' x_3' + x_1 x_2 x_3'^2 = 0.$$

7. **Ruled surfaces with common plane multiple curve $\overline{C}$.** Since any plane sections C, C' are in $(k, 1)$, $(k', 1)$ correspondence with $\overline{C}$ respectively, the equations of C and of C' may be written in the forms $f(\phi_1, \phi_2, \phi_3) = 0$, $f(\phi_1', \phi_2', \phi_3') = 0$, where the equations

$$y_i = \phi_i(x_1, x_2, x_3) = 0, \qquad y_i' = \phi_i'(x_1', x_2', x_3')$$

define the $(k, 1)$, $(k', 1)$ correspondences respectively. Any cases of this type can now be expressed by

$$f(\phi_1, \phi_2, \phi_3) = 0, \qquad f(\phi_1', \phi_2', \phi_3') = 0, \qquad \begin{vmatrix} A & B & C \\ \phi_1 & \phi_2 & \phi_3 \\ \phi_1' & \phi_2' & \phi_3' \end{vmatrix} = 0.$$

To obtain a fourth equation the procedure is exactly the same as in the preceding example.

8. **Ruled surfaces whose generators are multiple secants of a common curve $\overline{C}$.** A curve $\overline{C}$ can be found on R having the generators of R for multiple secants of any order. Another ruled surface R' can be constructed having for generators the bisecants of $\overline{C}$ which meet a given curve; still another may be found by the trisecants of $\overline{C}$.

Example: The bisecants of a space quartic $\overline{C}$ of genus 1 which meet a fixed line lie on a ruled surface R of order 8, having the line double, and $\overline{C}$ to multiplicity three. Let $\overline{C}$ be the intersection of the quadrics $\sum x_i^2 = 0$ $\sum a_i x_i^2 = 0$ and the line be $x_1 = 0$, $x_2 = 0$. This line passes through two of the vertices of the self-polar tetrahedron with respect to the pencil of quadrics through $\overline{C}$. The ruled surface is composite, consisting of a quartic R and of the two quadric cones with vertices $(0, 0, 0, 1)$, $(0, 0, 1, 0)$ and passing through $\overline{C}$. The two generators of R through a point $(0, 0, y_3, y_4)$ are the intersection of the quadric of the pencil through the point

$$(y_3^2 + y_4^2)(a_1 x_1^2 + a_2 x_2^2) - (a_3 y_3^2 + a_4 y_4^2)(x_1^2 + x_2^2)$$
$$+ (a_3 - a_4)(y_4^2 x_3^2 - y_3^2 x_4^2) = 0$$

with the tangent plane to the quadric at this point, thus

$$y_4 x_3 - y_3 x_4 = 0.$$

The equation of R is

$$(5)\qquad R \equiv (x_3^2 + x_4^2)(a_1 x_1^2 + a_2 x_2^2) - (a_3 x_3^2 + a_4 x_4^2)(x_1^2 + x_2^2) = 0.$$

Similarly, the equation of a ruled surface R' whose generators are the bisecants of $\overline{C}$ which meet $x_1 = 0$, $x_3 = 0$ is

$$(6)\qquad R' \equiv (x_2^2 + x_4^2)(a_1 x_1^2 + a_3 x_3^2) - (a_2 x_2^2 + a_4 x_4^2)(x_1^2 + x_3^2) = 0.$$

For C and C' we take the sections of these surfaces by the plane $x_2 = x_3$. Their equations are

$$(7)\qquad C \equiv (x_2^2 + x_4^2)(a_1 x_1^2 + a_2 x_2^2) - (x_1^2 + x_2^2)(a_3 x_2^2 + a_4 x_4^2) = 0,$$

$$(8)\qquad C' \equiv (x_2'^2 + x_4'^2)(a_1 x_1'^2 + a_3 x_2'^2) - (x_1'^2 + x_2'^2)(a_2 x_2'^2 + a_4 x_4'^2) = 0.$$

We shall now use y_i for current coördinates. The generator of R through (x) is given by

$$(9)\qquad x_2 y_1 = x_1 y_2, \qquad x_4 y_3 = x_2 y_4,$$

and the generator of R' through (x') by

$$(10)\qquad x_2' y_1 = x_1' y_3, \qquad x_4' y_2 = x_2' y_4.$$

The condition that these two generators intersect is

$$(11)\qquad \Delta \equiv x_2^2 x_1' x_4' - x_2'^2 x_1 x_4 = 0.$$

This relation, however, is not sufficient to determine the correspondence. Two of the intersections of (9) and (6) lie on $\overline{C}$ and two on K. The two intersections on C lie on the two quadrics and are therefore given by

$$(12)\qquad -\frac{y_2^2}{y_3^2} = \frac{x_2^2 + x_4^2}{x_1^2 + x_2^2} = \frac{a_3 x_2^2 + a_4 x_4^2}{a_1 x_1^2 + a_2 x_2^2}.$$

Hence (11) and

$$(13)\qquad x_2^2 x_1'^2 (x_1^2 + x_2^2) = x_1^2 x_2'^2 (x_2^2 + x_4^2) = 0$$

determine the correspondence associated with $\overline{C}$.

The four intersections of (9) and (6) are given by

$$(x_2^2 y_2^2 + x_4^2 y_3^2)(a_1 x_1^2 y_2^2 + a_3 x_2^2 y_3^2) - (x_1^2 y_2^2 + x_2^2 y_3^2)(a_2 x_2^2 y_2^2 + a_4 x_4^2 y_3^2) = 0;$$

hence the two intersections on K are given by

$$-\frac{y_2^2}{y_3^2} = \frac{(a_3 - a_4) x_4^2 (x_1^2 + x_2^2)}{(a_1 - a_2) x_1^2 (x_2^2 + x_4^2)}.$$

Hence (11) and

$$x_2^2 x_1'^2 (a_1 - a_2) x_1^2 (x_2^2 + x_4^2) = x_1^2 x_2'^2 (a_3 - a_4) x_4^2 (x_1^2 + x_2^2)$$

determine the correspondence associated with K.

9. **Cases of $(2, 2)$ correspondences.** If the curve K has two $(1, 2)$ involutions, there is a $(1, 1)$ transformation associated with each. The product of these two transformations must be of finite order if P is greater than 1. A repetition of the correspondence from curve to curve can therefore give only a finite number of images on each curve.

CASE OF $P = 1$. An elliptic curve with periods 2ω, $2\omega'$, has three irrational involutions of order 2 which transform K into elliptic curves C_1, C_2, C_3 of periods $(\omega, 2\omega')$, $(2\omega, \omega')$, $(2\omega, \omega + \omega')$.*

Between any two of these curves C a $(2, 2)$ correspondence exists, not definable by one equation. A repetition of the process, however, reduces all these curves back to K, so that the $(2, 2)$ correspondences are each compounded of two $(1, 2)$ involutions. Using non-homogeneous coördinates, each involution may be expressed by means of a $(1, 2)$ correspondence between the planes (x', y') and (x, y), where K is a cubic of the form

$$y^2 = 4(x - e_1)(x - e_2)(x - e_3).$$

If we now put $x = \wp(u)$, $y = \wp'(u)$, then the relation between $K(x)$ and $C_1(x')$ is expressed by the equations

$$x' = x + \frac{(e_2 - e_1)(e_3 - e_1)}{x - e_1}, \qquad y' = y\left[1 - \frac{(e_2 - e_1)(e_3 - e_1)}{(x - e_1)^2}\right];$$

similar forms exist for C_2 and C_3.

If we indicate by $K(u)$ the point on K having the parameter u, we may say: given a point $K(u)$, there exists one image point $C_i(u)$ on each curve C_i. The residual image of $C_1(u)$ is $K(u + \omega)$, of $C_2(u)$ $K(u + \omega')$, of $C_3(u)$ $K(u + \omega + \omega')$. These four points on K form a closed set. They all have the point $K(-2u)$ for first tangential.

CASE OF $P = 3$. The quartic K of genus 3

$$a(x_1^4 + x_2^4) + bx_1^2 x_2^2 + cx_1 x_2 x_3^2 + dx_3^4 = 0$$

possesses the two $(1, 2)$ involutions

$$x_1' = (x_1 + x_2)x_3, \qquad x_2' = x_1 x_2, \qquad x_3' = x_3^2,$$

$$x_1'' = (ix_1 + i^3 x_2)x_3, \qquad x_2'' = x_1 x_2, \qquad x_3'' = x_3^2$$

which transform K into elliptic quartics C_1, C_2. The curve K is invariant under the corresponding $(1, 1)$ transformations

$$x_1' = x_2, \qquad x_2' = x_1, \qquad x_3' = x_3$$

and

$$x_1' = ix_1, \qquad x_2' = i^3 x_2, \qquad x_3' = x_3.$$

The product of these two transformations is of period 2.

* Bianchi, *Lezioni sulla teoria delle funzioni ellittiche*, Seconda edizione. See p. 485.

The $(2,2)$ correspondence between $C_1(x')$ and $C_2(x'')$ is defined by the two equations

$$4x_2'x_3'x_2''^2 = x_1'^2x_2''^2 + x_1''^2x_2'^2, \qquad x_2'x_3'' = x_2''x_3'.$$

A GENERAL CASE. This case can be generalized by taking for K an equation of the form of a polynomial in $x_1^n + x_2^n, x_1x_2, x_3$ equated to zero, and replacing i by θ where $\theta^n = 1$. We thus find $2n$ points on K and n points on each curve C which form closed sets. The equations of the $(2,2)$ correspondence between $C(x)$ and $C'(x')$ are

$$x_1^2x_3'^2 - x_1x_1'x_3x_3'(\theta+\theta^{-1}) + x_1'^2x_3 + x_2x_3x_3'^2(\theta-\theta^{-1})^2 = 0,$$

$$x_2x_3' - x_3x_2' = 0.$$

The point $(1,0,0)$ is not on C nor on C'. The genus of K is $(n-1)(n-2)/2$; hence the $(2,2)$ correspondence is closed. The product of the two linear transformations which leave K invariant is of order 2; the group generated by them is dihedral of order $2n$.

10. **General case. Analytical form of** K. Since K possesses two involutions, one of order n, the other of order n', it follows that if the equations of C and C' are $f(x) = 0, f'(x') = 0$, the equation of K may be written in the two forms $f(\phi(y)) = 0, f'(\phi'(y')) = 0$, where $x_i = \phi_i(y), x_1' = \phi_1'(y')$ define a $(1,n)$ and $(1,n')$ transformation respectively. If no restrictions are put on the coefficients in f and ϕ_i it is impossible to write K in the form $f'(\phi'(y')) = 0$. We may, however seek to find the restrictions on f and ϕ so that the second form is possible.

Example: If $f(x) = 0$ is a general cubic and each ϕ is a general quadratic in (y) we proceed to determine restrictions on the coefficients in f and each ϕ so that we may write K in the form

$$(y_1^3 + y_1^2f_1 + y_1f_2 + f_3)^2 = f_6, \tag{14}$$

where f_i is a binary form of order i in y_2 and y_3. By linear transformations on (x) and (y) we may reduce f_1 to zero and take

$$\begin{aligned} x_1 &= \phi_1 = y_1^2 + ay_2^2 + cy_3^2, \\ x_2 &= \phi_2 = y_1(y_2+y_3) + dy_2^2 + ey_3^2, \\ x_3 &= \phi_3 = y_2y_3. \end{aligned} \tag{15}$$

The form of f is then

$$\begin{aligned} &x_1^3 + x_1^2x_3 + x_1(Ax_2^2 + Bx_2x_3 + Cx_3^2) \\ &\qquad + D^3x_2^3 + Ex_2^2x_3 + Fx_2x_3^2 + Gx_3^3 = 0. \end{aligned} \tag{16}$$

The restrictions on the coefficients are much simplified by taking $A = 0$,

but this is not necessary. If we substitute $x_i = \phi_i$ from (15) in (16) and make the result identical with (14) we find the solution

$$a = c = 2D, \qquad d = e = -D,$$

$$A = 0, \qquad B = \frac{D}{2} - 6D^3, \qquad C = \frac{1}{4} + D, \qquad E - \frac{D^2}{2}, \qquad F = \frac{D}{4} - 2D^3,$$

$$f_2 = 3D^2(y_2^2 + y_3^2) + \frac{y_2 y_3}{2}, \qquad f_3 = \frac{B}{2} y_2 y_3 (y_2 + y_3) + \frac{D^3}{2}(y_2 + y_3)^3.$$

The value of G is arbitrary; the form of f_6 follows from the preceding values of the other quantities involved. The transformation

$$x_1' = y_1^3 + y_1 f_2 + f_3,$$
$$(17) \qquad x_2' = y_2 y_3^2,$$
$$x_3' = y_3^3$$

reduces K to the sextic of genus 2

$$(18) \qquad x_1'^2 x_3'^3 = f_6(x_2', x_3').$$

Between the curves (16) and (18) exists a $(3, 4)$ correspondence defined by the $(1, 3)$ correspondence (17) and the $(1, 4)$ correspondence defined by (15). If now we eliminate (y) between (15) and (17) we obtain the two equations of the correspondence.

11. **Surfaces defined by pairs of points of two algebraic curves.** An important method of representing multiple correspondences between two curves is that of a surface Σ such that to a point P on the surface corresponds a point P_1 on one curve and a point P_2 on the other.*

Let the curves be defined by

$$(19) \qquad f(x_1, x_2, x_3) = 0, \qquad x_1' = 0, \qquad x_2' = 0,$$

$$(20) \qquad f'(x_1', x_2', x_3') = 0, \qquad x_1 = 0, \qquad x_2 = 0$$

in the space of four dimensions $S_4 \equiv (x_1, x_2, x_3, x_1', x_2', x_3')$, $x_3 = x_3'$; and let Σ be defined by $f(x_1, x_2, x_3) = 0, f'(x_1', x_2', x_3') = 0$. A plane $k_3 x_1 = k_1 x_3$, $k_3 x_2 = k_2 x_3$ belonging to the conical variety $f = 0$ meets Σ in the plane curve

* F. Severi, *Sulle corrispondenze fra i punti di una curva algebrica e sopra certe classi di superficie*, Memorie della Accademia reale di Torino, vol. 54 (1903), pp. 1–49. See pp. 19–34.

$$f'(x'_1, x'_2, x'_3) = 0, \qquad k_3 x_1 - k_1 x_3 = 0, \qquad k_3 x_2 - k_2 x_3 = 0.$$

Thus, Σ contains a one-dimensional system of plane curves, each birationally equivalent to the curve (20). Similarly, by intersecting the variety $f = 0$ by the planes belonging to $f' = 0$, we obtain a second system of plane curves, each birationally equivalent to the curve (19). Since two planes in S_4 meet in one point, it follows that every curve of each system meets every curve of the other in one and only one point.

12. **Multiple correspondences on Σ.** Let K be any algebraic curve on Σ. It meets the curves of one system in n points, and those of the other in n' points. Since through every point of K one curve of each system passes, it follows that K establishes an (n, n') correspondence between the curves (19) and (20). Conversely, any correspondence between these curves may be represented by a curve K on Σ.

When the correspondence can be defined by one equation, K is a complete intersection on Σ, and conversely. When the correspondence requires two equations for its definition, K is a partial intersection on Σ. The results already found by means of ruled surfaces can be readily interpreted in terms of K on Σ. Thus, when two ruled surfaces have a common generator, their residual intersection corresponds to a curve K on Σ which has a double point for each common generator. If two ruled surfaces have a common simple directrix this curve corresponds to the curve of the identical transformation on Σ and $\Delta = 0$ is a variety through this directrix, meeting Σ in a residual curve K. Similarly for a multiple directrix.

Moreover, we see that if between two curves exists one multiple correspondence not expressible by means of one equation, then these curves also have other such correspondences formed by passing a variety through the curve of the given correspondence on Σ, and taking the residual intersection.

13. **General criteria.** The problem of finding correspondences between two curves that require two equations for their determination may be presented in a different form by commencing with $f(x) = 0$, $\phi(x, x') = 0$, in which the coefficients in each are as yet undetermined. We then find the restrictions on these functions so that sets of points in (x) determined by $f = 0$, $\phi = 0$ are rationally separable into two or more sets for values of the x'_i which satisfy an equation $f'(x') = 0$.

Example. Let $f(x) = 0$, $\phi(x, x') = 0$ be respectively

$$x_1^4 + x_2^4 + x_3^4 = 0, \qquad x_1^2 x_1'^2 + x_2^2 x_2'^2 + x_3^2 x_3'^2 = 0.$$

Eliminate x_1 between these two equations and express the condition that the resultant quadratic in x_2, x_3 is rationally factorable. This requires that the

expression $x_1'^4 + x_2'^4 + x_3'^4$ shall be a square for points on $f'(x') = 0$. We may put, for example,

$$f'(x') \equiv (x_1'^4 + x_2'^4 + x_3'^4)x_1'^4 = x_2'^6 x_3'^2.$$

Then

$$\phi_1(x, x') = (x_1'^4 + x_2'^4)x_2^2 + x_2 x_3 x_2'^2 x_3'^2 + x_2'^3 x_3' x_3^2 = 0.$$

The correspondence between the planes (x), (x') is $(4, 8)$. The genus of $f(x) = 0$ is 3, and of $f'(x') = 0$ is 9.

Cornell University

CONCERNING CERTAIN EQUICONTINUOUS SYSTEMS OF CURVES*

BY

ROBERT L. MOORE

In order that a system G of open curves lying in a given plane S should be equivalent, from the standpoint of analysis situs,† to a complete† system of parallel lines in S it is not sufficient that through each point of S there should

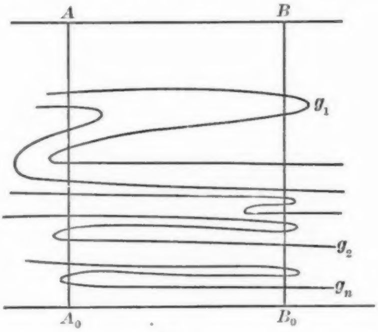

FIG. 1.

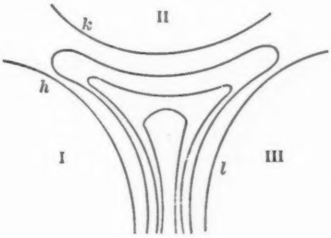

FIG. 2.

pass one and only one curve of the system G. Consider the examples indicated in Figs. 1 and 2.‡ In each of these examples through each point of the plane there is one and only one curve of the system in question but the system

* Cf. papers presented to the Society, April 28 and October 27, 1917.

† A *complete system* of parallel lines in a plane S is the set of all lines in S parallel to a given line. A system G of open curves is said to be equivalent, from the standpoint of analysis situs, to such a system of lines L if there is a one to one continuous transformation of S into itself which carries G into L.

‡ In the case roughly indicated by Fig. 1, $A_0 B_0$ and AB are two parallel lines at a distance apart equal to 1. These lines both belong to the system G and so does every line which is parallel to them but which does not lie between them. For each positive integer n, g_n is an open curve belonging to G such that (1) there is an interval of g_n that contains a point of $B_0 B$ but has its endpoints on $A_0 A$, (2) every point of g_n is at a distance of less than $1/n$ from the line $A_0 B_0$. Of course, as is indicated in Fig. 1 for the case $n = 1$, there does not exist, on *every* curve of G that lies between g_n and g_{n+1}, an interval that contains a point of $B_0 B$ and has its endpoints on $A_0 A$.

In the example indicated in Fig. 2, the open curves h, k and l belong to G. To obtain the curves of G which lie in domain I, II or III construct through each point of that domain an open curve parallel and congruent to h, k or l respectively. Each curve of G that lies in the domain bounded by h, k and l lies as is roughly suggested in the figure.

is not in one to one continuous correspondence with a complete system of parallel lines. Let G_1 and G_2 be the system of curves represented in Figs. 1 and 2 respectively. The system G_1 is not equicontinuous.* That is to say it is not true that for every positive number ϵ there exists a positive number δ_ϵ such that if P_1 and P_2 are points on some curve g of G at a distance apart less than δ_ϵ then that arc of g which has P_1 and P_2 as its endpoints lies wholly within some circle of radius ϵ. The system G_2 is equicontinuous but fails to be what I will call inversely equicontinuous.

Definition 1. A system of curves G is *equicontinuous with respect to a given point-set M* if for every positive number ϵ there exists a positive number $\delta_{M\epsilon}$ such that if P_1 and P_2 are two points of M at a distance apart less than $\delta_{M\epsilon}$ and lying on a curve g of the system G then that arc of g which has P_1 and P_2 as endpoints lies wholly within some circle of radius ϵ.

Definition 2. A system of curves G is *inversely equicontinuous with respect to a point-set M* if for every positive number ϵ there exists a positive number $\delta_{M\epsilon}$ such that if P_1 and P_2 are two points of M at a distance apart less than ϵ and lying on a curve g of the system G then that interval of g which has P_1 and P_2 as endpoints lies wholly within a circle of radius $\delta_{M\epsilon}$.

I will show that if G is a system of open curves lying in S such that through each point of S there is just one curve of G, then in order that the system G should be equivalent, from the standpoint of analysis situs, to a complete system of parallel straight lines it is necessary and sufficient that it should be both equicontinuous and inversely equicontinuous with respect to every bounded set of points. Additional theorems of a related nature will also be established.

Theorem 1. *Suppose that, in a given plane S, $ABCD$ is a rectangle and G is a set of arcs such that* (1) *through each point of the point-set $\overline{R}$, composed of $ABCD$ and its interior R, there is just one arc of G,* (2) *BC and AD are arcs of G,* (3) *every arc of G (with the exception of BC and AD) lies entirely within $ABCD$ except that its endpoints are on AB and CD respectively,* (4) *the set of arcs G is equicontinuous.*

Then there is a one to one continuous transformation of the plane S into itself which transforms the rectangle $ABCD$ into a rectangle $A'B'C'D'$ and transforms the set of arcs G into the set of all straight line intervals which are parallel to $A'D'$ and lie between $A'D'$ and $B'C'$ (except that one of them coincides with $A'D'$ and another with $B'C'$) and are terminated by $A'B'$ and $C'D'$.

The truth of this theorem will be established with the help of a lemma. This lemma will be proved first.

Definition 3. A connected domain K is said to be a *simple domain with*

* Cf. G. Ascoli, *Sulle curve limiti di una varietà data di curve*, Memorie della Reale Accademia dei Lincei, vol. 18 (1884), pp. 521–586.

respect to a set of arcs G satisfying the conditions stated in the hypothesis of Theorem 1 if (1) every point of K is within $ABCD$, (2) K contains the whole of every G-interval* whose endpoints are in K, (3) there exist two G-arcs g_1 and g_2 such that (a) g_1 lies above g_2, every point of K is between g_1 and g_2 and both g_1 and g_2 have points in common with the boundary of K, (b) the set of all those points that the boundary of K has in common with g_i is an interval t_i of g_i $(i = 1, 2)$, (c) no point of t_1 or of t_2 is a limit point of a point-set which lies between g_1 and g_2 and contains no point of K. The interval t_i minus its endpoints will be called the upper base, and the interval t_2 minus its endpoints will be called the lower base, of the domain K.

LEMMA 1. *If G is a set of arcs satisfying the conditions stated in the hypothesis of Theorem* 1 *and K is a simple domain with respect to G, then any point on the upper base of K can be joined to any point on its lower base by a simple continuous arc that lies wholly in K and does not have more than one point in common with any arc of the set G.*

Proof. If P is a point of R and ϵ is a positive number let $R_{P\epsilon}$ denote the set of all points X such that X lies on a G-interval whose endpoints are both within a circle of radius ϵ with center at P. If for a given point P and a given pair of positive numbers e and ϵ, such that $e \leqq \epsilon$, the point-set R_{Pe} has points between two distinct G-arcs g_1 and g_2 and also has points on g_1 and points on g_2, the set of all those points of R_{Pe} that lie between g_1 and g_2 will be called an elemental region of rank ϵ.† It may be easily proved that if ϵ is a positive number each point of K is in some elemental region of rank ϵ which lies together with its boundary wholly in the point-set K^* composed of K and its two bases. Such an elemental region will be called a K-element of rank ϵ. If E and F are two points of K^* and E is above F, a *chain of K-elements from E to F* or *from F to E* or *joining E to F* or *F to E* is a finite set of K-elements $K_1, K_2, K_3, \cdots, K_n$ such that (1) E belongs to the upper base of K_1 and F belongs to the lower base of K_n, (2) for each i $(1 \leqq i \leqq n)$ the lower base of K_i and the upper base of K_{i+1} lie on the same arc of the set G and have points in common and the set of all their common points is a segment t_i. The point-set $K_1 + K_2 + K_3 + \cdots + K_n + t_1 + t_2 + t_3 + \cdots + t_{n-1}$ is a simple domain. It will be called the domain associated with the chain $K_1, K_2, \cdots, K_n$. Suppose that E is a point on the upper base of K, F is a point on its lower base and ϵ is a positive number. I will show that E can be joined to F by a chain of K-elements of rank ϵ. Let $\overline{K}$ denote the set of all those points of K

* If G is a set of arcs or curves a *G-arc* or a *G-curve* is an arc or a curve of the set G. A *G-interval* is an interval (and a G-segment is a segment) of such an arc or curve. If G is a set satisfying the conditions stated in the hypothesis of Theorem 1, the G-arc g_1 is said to be *above* the G-arc g_2 if it lies between g_2 and BC. If P is a point of $\overline{R}$, g_P denotes that arc of G which contains P. If P_1 and P_2 are points of $\overline{R}$, P_1 will be said to lie above P_2 in case g_{P_1} is above g_{P_2}.

† According to this definition if $\epsilon_1 < \epsilon_2$ every elemental region of rank ϵ_1 is also of rank ϵ_2.

that lie on arcs of G below the arc g_E and that can be joined to E by chains of K-elements of rank ϵ. There exists a K-element of rank ϵ whose upper base contains the point E and every such K-element contains points in common with some g-arc lying below g_E. It follows that the set $\overline{K}$ exists.

Suppose that WZ is an arc of G that contains a point of $\overline{K}$. The set of points common to WZ and K is a segment $W'Z'$. Every point of $W'Z'$ must belong to $\overline{K}$. For suppose this is not the case. Then the segment $W'Z'$ is the sum of two mutually exclusive point-sets S_1 and S_2 such that S_1 is a subset of $\overline{K}$ but no point of S_2 belongs to $\overline{K}$. There exists a point P which either belongs to S_1 and is a limit point of S_2 or belongs to S_2 and is a limit point of S_1. In the first case there is a chain α_2 of K-elements of rank ϵ from E to P. The lower base of the last element of this chain is a segment of $W'Z'$ containing P. Since P is a limit point of S_2 this segment must contain at least one point P_2 of S_2. Thus α_2 is a chain of K-elements of rank ϵ from E to P_2. Thus the supposition that S_1 contains a limit point of S_2 leads to a contradiction. Suppose now that S_2 contains a point P which is a limit point of S_1. There exists (Fig. 3) a K-element e of rank ϵ whose lower base

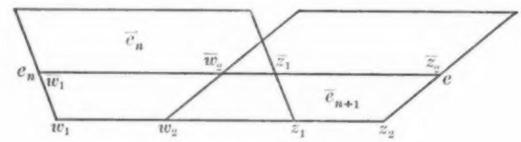

FIG. 3.

W_2Z_2 is a segment of $W'Z'$ containing P. Since P is a limit point of S_1 there exists on the segment W_2Z_2 a point P_1 belonging to S_1. There exists a chain $e_1, e_2, e_3, \cdots, e_n$ of K-elements of rank ϵ from E to P_1. The lower base of the last element e_n of this chain is a segment W_1Z_1 containing P_1. There exist a G-arc $\bar{g}$ and two segments $\overline{W}_1\overline{Z}_1$ and $\overline{W}_2\overline{Z}_2$ such that (1) $\overline{W}_1\overline{Z}_1$ is the set of all points common to e_n and $\bar{g}$, (2) $\overline{W}_2\overline{Z}_2$ is the set of all points common to e and $\bar{g}$, (3) $\overline{W}_1\overline{Z}_1$ and $\overline{W}_2\overline{Z}_2$ have a segment in common. Let $\bar{e}_n$ denote that part of e_n which lies between $\bar{g}$ and the arc of G that contains the upper base of e_n. Let $\bar{e}_{n+1}$ denote that part of e which lies between $\bar{g}$ and W_2Z_2. The set of elements $e_1, e_2, e_3, \cdots, e_{n-1}, \bar{e}_n, \bar{e}_{n+1}$ is a chain of K-elements of rank ϵ from E to P. It is thus established that if one point of $W'Z'$ belongs to $\overline{K}$ then so does every other point of $W'Z'$. It has been shown that if a G-arc above g_F contains a point of $\overline{K}$ then so must some lower arc of G. It follows that if F does not belong to $\overline{K}$ there exists an arc XY which is the uppermost arc of G that contains no point of $\overline{K}$. Let P denote a point of K on the arc XY. There exists a K-element e of rank ϵ whose lower base contains P. The set G contains an arc g that intersects e in a segment MN.

Let $\bar{P}$ denote a point of MN. There exists a chain of K-elements of rank ϵ from E to $\bar{P}$. If to this chain of elements there is added that portion of the K-element e which lies between g and XY there is obtained a chain of K-elements of rank ϵ from E to P. Thus the supposition that E can not be joined to F by a chain of K-elements of rank ϵ leads to a contradiction. It follows that there exists a simple chain $e_{11}, e_{12}, e_{13}, \cdots, E_{1n}$ of K-elements of rank 1 from E to F. Let K_1 denote the domain associated with this chain. There exists a simple chain of K_1-elements of rank 1 from E to F. This process may be continued. It follows that there exists a sequence of simple chains $C_1, C_2, C_3, \cdots$ from E to F such that if, for each n, K_n denotes the domain associated with C_n then (1) every link of C_{n+1} is a K_n-element of rank $1/n$, (2) K'_{n+1} is a subset of the point-set composed of K_n plus its bases. Let t denote the set of all points $[X]$ such that X belongs to every K_n. With the aid of the fact that the set G is equicontinuous, it can be proved* that t is a simple continuous arc from E to F and that it does not have more than one point in common with any given arc of the set G. The truth of Lemma 1 is thus established.

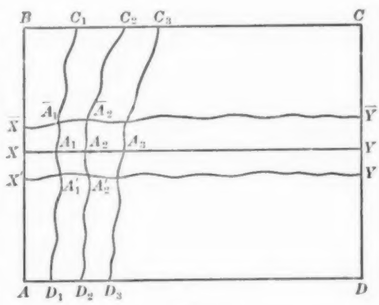

FIG. 4.

Proof of Theorem 1. If X is a point of AB and XY is that arc of G which has X as one of its endpoints, it may be easily proved with the aid of the Heine-Borel Theorem that there exists on XY a finite set of points $A_1, A_2, A_3, \cdots, A_n$ in the order $XA_1 A_2 A_3 A_4 \cdots A_{n-1} A_n Y$ such that each of the intervals $XA_1, A_1 A_2, \cdots, A_{n-1} A_n, A_n Y$ of the arc XY lies wholly within some circle of radius 1. Let $C_1, C_2, C_3, \cdots, C_n$ denote n points in the order $BC_1 C_2 C_3 \cdots C_{n-1} C_n C$ on the arc BC and let $D_1, D_2, D_3, \cdots, D_n$ denote n points in the order $AD_1 D_2 \cdots D_{n-1} D_n D$ on the arc AD. With the use of Lemma 1 it is easily established that there exist (Fig. 4) two sets of arcs $A_1 C_1, A_2 C_2, A_3 C_3, \cdots, A_n C_n$ and $A_1 D_1, A_2 D_2, A_3 D_3, \cdots, A_n D_n$ such that no arc of either set has a point in common with any other arc of that set and such that, for every n, (1) $A_n C_n$ lies except for its endpoints entirely within $ABCD$ and between XY and BC, (2) $A_n D_n$ lies, except for its endpoints, entirely within $ABCD$ and between XY and AD, (3) neither $A_n C_n$ nor $A_n D_n$ has more than one point in common with any one arc of the set G. It is easy to show that there exist two points X' and $\bar{X}$ in the order $AX' X\bar{X}B$

* Cf. the proof of Theorem 15 of my paper *On the foundations of plane analysis situs*, these Transactions, vol. 17 (1916), pp. 136–139.

and two arcs $X'Y'$ and $\bar{X}\bar{Y}$ belonging to G such that if for every i $(1 \leqq i \leqq n)$ A_i' is the point in which $X'Y'$ intersects A_iD_i and $\bar{A}_i$ is the point in which $\bar{X}\bar{Y}$ intersects A_iC_i then the closed curve bounded by the intervals $A_i'A_i$, $A_i\bar{A}_i$, $\bar{A}_i\bar{A}_{i+1}$, $\bar{A}_{i+1}A_{i+1}$, $A_{i+1}A_{i+1}'$ and $A_{i+1}'A_i'$ of the arcs A_iD_i, A_iC_i, $\bar{X}\bar{Y}$, $A_{i+1}C_{i+1}$, $A_{i+1}D_{i+1}$ and $X'Y'$ respectively (Fig. 4) lies entirely within some circle of radius 1. For each point X of AB make a similar construction and apply the Heine-Borel Theorem to the set of segments $[X'\bar{X}]$. If certain arcs are properly continued there will result a double ruling* T_1 of $ABCD$ such that (1) the arcs of one of its single rulings are arcs of G and each arc of its other single ruling has its endpoints on BC and AD respectively and has just one point in common with each arc of the set G, (2) each of the subdivisions into which T_1 divides $ABCD$ lies within some circle of radius 1. In a similar way each subdivision α of this set can itself be subdivided by a double ruling $T_{1\alpha}$ such that (1) each arc of one of its single rulings is an interval of an arc of G, (2) each arc of its other single ruling has its endpoints on the arcs which form respectively the upper and the lower base of α and no arc of this ruling has more than one point in common with any arc of G, (3) each of the subdivisions into which $T_{1\alpha}$ divides α is within a circle of radius $1/2$. It follows that there exists a double ruling T_2 satisfying the Conditions (1) and (2) stated above as being satisfied by T_1 and also satisfying the additional condition that each of its subdivisions is within some circle of radius $1/2$, for every α each arc of $T_{1\alpha}$ being an interval of an arc of one or the other of the rulings of T_2. This may be continued. It follows that there exists an infinite sequence of double rulings $T_1, T_2, T_3, \cdots$ such that for every n, (1) T_n satisfies the conditions (1) and (2) stated above for T_1, (2) each arc of T_n is an arc of T_{n+1}, (3) each subdivision of T_n is within a circle of radius $1/n$. Let β be the set of all arcs $[t]$ such that, for some n, t belongs to one of the rulings of T_n and has its endpoints on AD and BC respectively. If P is a point on BC which is not an endpoint of an arc of the set β then there exists just one arc t_P that has one endpoint at P and the other on AD, lies except for its endpoints entirely within $ABCD$ and has no point in common with any arc of the set β. Let γ be the set of all such arcs t_P for all such points P. Let G' denote the set of arcs composed of all the arcs of β together with all the arcs of γ and the straight intervals AB and CD. If P is a point on or within the rectangle $ABCD$ let h_P denote the distance from A to the point of intersection of AD with that arc of G' that passes through P. Let k_P denote the distance from A to the point in which AB intersects that arc of G which passes through P. Let AD be the axis of X and AB the axis of Y in a rectangular system of coördinates. If P is on or within the rectangle $ABCD$ let

* Cf. my paper *Concerning a set of postulates for plane analysis situs*, these Transactions, vol. 20 (1919), p. 172 (footnote) and pp. 172–175.

P' denote the point whose coördinates are (h_P, k_P). Let $\overline{T}$ denote the transformation of $\overline{R}$ into itself such that if P is any point of $\overline{R}$ then $\overline{T}(P) = P'$. It is easy to see that the transformation $\overline{T}$ is continuous and that there exists a continuous transformation T, of S into itself, which reduces to $\overline{T}$ on $\overline{R}$. The transformation T satisfies all the requirements of Theorem 1.

THEOREM 2. *If, in a plane S, G is a set of open curves such that through each point of S there is just one curve of G, then in order that the set of curves G should be in one to one continuous correspondence with a complete system of parallel lines in S it is necessary and sufficient that the set G should be both equicontinuous and inversely equicontinuous with respect to every bounded set of points.*

That this condition is necessary may be easily seen. I will show that it is sufficient.

Proof. Suppose that G is a set of open curves such that (1) through each point of S there is just one curve of G, (2) G is both equicontinuous and inversely equicontinuous with respect to every bounded set of points. I will first show that of any three distinct curves of the set G one separates the other two from each other.

Suppose on the contrary that there exist three open curves h, k and l of the set G such that no one of them separates the other two. Then the set of all points $[P]$ such that P is between every two of the curves h, k and l is a domain D. Every curve of G which contains a point of D lies wholly in D. If g is any curve of G lying wholly in D then either (1) g separates one of the curves h, k and l from the other two or (2) two of the curves h, k and l are such that if they be designated as $\overline{h}$ and $\overline{k}$ respectively and the third one be designated as $\overline{l}$ then there exists a ray AB of $\overline{h}$, a ray CD of $\overline{k}$ and an arc AC lying except for its endpoints wholly in D such that the rays AB and CD and the arc AC constitute the common boundary of a domain E which contains g and is a subset of D. A curve g satisfying condition (1) will be called a curve of class I with respect to that one of the curves h, k and l which it separates from the other two, and a curve satisfying condition (2) will be called a curve of class II with respect to $\overline{h}$ and $\overline{k}$. Suppose there exist curves of class I with respect to h. It is clear that of every two such curves one of them separates the other one from h and is separated by the other one from k and from l. I will show that there is a last curve of class I with respect to h, that is to say there is one that separates every other one from k and from l. Suppose this is not the case. Let K denote a point of k and L a point of l and let KL denote an arc which lies except for its endpoints entirely in D. In view of the fact that the set of curves G is inversely equicontinuous with respect to the bounded point-set KL, it is clear that there exist two other points K' and L' on k and l respectively and an arc $K'\,L'$ lying, except for its endpoints,

wholly in D and having no point in common with KL such that (1) the rays $K'K$ and $L'L$ of k and l respectively, together with the arc $K'L'$ and the curve h, constitute the complete boundary of a domain α which is a subset of D and (2) no arc of G with endpoints on KL contains a point of $K'L'$. There must exist a point-set β (Fig. 5) which is a subset of α + ray $K'K$

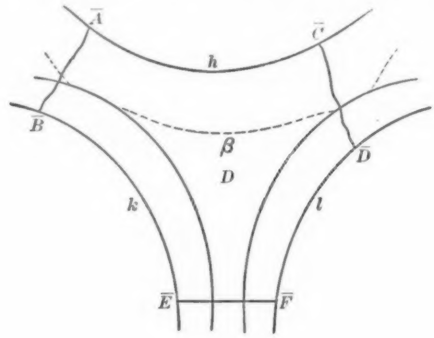

Fig. 5.

+ ray $L'L$ such that β is the complete boundary of the set of all points $[X]$ such that X is separated from k and from l by some curve of class I with respect to h. The point-set β is connected and contains points in α. For each such point P there exists through P a curve g_P of the set G. Let $\bar{g}_P$ denote the set of all those points that are common to g_P and β. The point-set $\bar{g}_P$ is a closed proper subset of the connected point-set β and has no point in common with k or with l. It follows that if $\bar{P}$ is a definite point of β lying in α the point-set $\bar{g}_{\bar{P}}$ contains a point P_0 which is the sequential limit point of a sequence of points $P_1, P_2, P_3, \cdots$, all belonging to β and lying in α such that (1) no two of the point-sets $\bar{g}_{P_1}, \bar{g}_{P_2}, \bar{g}_{P_3}, \cdots$ lie on the same curve of the set G and (2) the point-set $P_1 + P_2 + P_3 + \cdots$ is within some closed curve J that lies wholly in D. There exist six distinct points $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{D}$, $\bar{E}$ and $\bar{F}$ and arcs $\bar{A}\bar{B}$, $\bar{C}\bar{D}$, $\bar{E}\bar{F}$ such that (1) $\bar{A}$ and $\bar{C}$ are on h, $\bar{B}$ and $\bar{E}$ are on k, and $\bar{D}$ and $\bar{F}$ are on l, (2) each of the arcs $\bar{A}\bar{B}$, $\bar{C}\bar{D}$ and $\bar{E}\bar{F}$ lies, except for its endpoints, in the domain D and no two of them have a point in common, (3) the curve J is wholly within the closed curve $\bar{J}$ formed by the arcs $\bar{A}\bar{B}$, $\bar{E}\bar{F}$ and $\bar{C}\bar{D}$ together with the intervals $\bar{A}\bar{C}$, $\bar{B}\bar{E}$ and $\bar{D}\bar{F}$ of the curves h, k and l respectively. There does not exist more than one integer n such that the curve g_{P_n} separates k from h and from l. For suppose there are two such integers n_1 and n_2. Then one of the curves $g_{P_{n_1}}$ and $g_{P_{n_2}}$ separates the other one from h and therefore separates a point of β from h. But every domain that contains a point of β contains a point of some G-curve that separates h from k and from l. Hence either $g_{P_{n_1}}$ or $g_{P_{n_2}}$ separates h

from a G-curve g_0 which separates h from k and from l. Hence $g_{P_{n_1}}$ or $g_{P_{n_2}}$ has at least one point in common with g_0. But this is contrary to hypothesis. Similarly there does not exist more than one value of n such that g_{P_n} separates l from h and from k. It follows that there exists an infinite sub-sequence $g_{P_{n_1}}, g_{P_{n_2}}, g_{P_{n_3}}, \cdots$ of distinct curves of the sequence $g_{P_1}, g_{P_2}, g_{P_3}, \cdots$ and an arc XY (identical with one of the arcs $\overline{AB}$, $\overline{CD}$ and $\overline{EF}$) such that for each m the curve $g_{P_{n_m}}$ contains an interval $A_{P_{n_m}} P_{n_m} B_{P_{n_m}}$ whose endpoints A_{P_n} and B_{P_n} are on XY and which lies, except for its endpoints, wholly within $\bar{J}$. By hypothesis, for every positive number ϵ there exists a positive number $\delta_{\bar{J}_\epsilon}$ such that if, for some n, the distance from $A_{P_{n_m}}$ to $B_{P_{n_m}}$ is less than $\delta_{\bar{J}_\epsilon}$ then the whole arc $A_{P_{n_m}} P_{n_m} B_{P_{n_m}}$ lies within a circle of radius ϵ. It can be easily seen that if i and j are distinct integers the intervals $A_{P_{n_i}} B_{P_{n_i}}$ and $A_{P_{n_j}} B_{P_{n_j}}$ of the arc XY have no point in common. Hence if ϵ is the least distance from a point of the arc XY to a point of the closed curve J there exists an integer $\overline{m}$ such that the distance from $A_{P_{n_{\overline{m}}}}$ to $B_{P_{n_{\overline{m}}}}$ is less than $\delta_{\bar{J}_\epsilon}$. It follows that every point of the arc $A_{P_{n_{\overline{m}}}} P_{n_{\overline{m}}} B_{P_{n_{\overline{m}}}}$ is at a distance of less than ϵ from the point $A_{P_{n_{\overline{m}}}}$. But the distance from $P_{n_{\overline{m}}}$ to $A_{P_{n_{\overline{m}}}}$ is not less than ϵ. Thus the supposition that there exists no last curve of class I with respect to h has led to a contradiction. Hence there exists a curve $\bar{h}$ which is the last curve of class I with respect to h. In a similar way it may be shown that there exist curves $\bar{k}$ and $\bar{l}$ which are the last curves of class I with respect to k and l respectively. No one of the curves $\bar{h}$, $\bar{k}$ and $\bar{l}$ separates the other two from each other and no curve of the set G separates one of them from the other two.

Let $\bar{D}$ denote the connected domain which is bounded by the curves $\bar{h}$, $\bar{k}$ and $\bar{l}$. With the aid of several applications of the fact that the system G is inversely equicontinuous with respect to every bounded set of points it can be shown that there exist six points $\overline{M}$, $\overline{T}$, $\overline{H}$, $\overline{L}$, $\overline{K}$, $\overline{N}$ and three arcs $\overline{MN}$, $\overline{TH}$ and $\overline{KL}$ such that (1) $\overline{M}$ and $\overline{T}$ are on $\bar{h}$, $\overline{H}$ and $\overline{L}$ are on $\bar{l}$, $\overline{K}$ and $\overline{N}$ are on $\bar{k}$, (2) $\overline{MN}$, $\overline{TH}$ and $\overline{KL}$ lie, except for their endpoints, entirely in $\bar{D}$ and no two of them have a point in common, (3) no curve of G distinct from $\bar{h}$, $\bar{k}$ and $\bar{l}$ contains a point of more than one of the arcs $\overline{MN}$, $\overline{TH}$ and $\overline{KL}$ (Fig. 6). Let $\bar{\beta}$ denote the region bounded by the arcs $\overline{MN}$, $\overline{TH}$, $\overline{KL}$ and the intervals $\overline{MT}$, $\overline{HL}$ and $\overline{KN}$ respectively of the curves $\bar{h}$, $\bar{l}$ and $\bar{k}$. By an argument similar in large part to that employed above to show the existence of $\bar{h}$ it may be proved that if g is a curve of the set G that contains a point of $\overline{MN}$ there exists a curve $\bar{g}$ of the set G which either coincides with g or separates g from each of the curves $\bar{h}$, $\bar{k}$ and $\bar{l}$ but is not itself separated from any one of these curves by any other curve of G. Every such curve $\bar{g}$ will be called a curve of class III. There clearly exist infinitely many distinct curves of class III. Let M^* denote a point on $\bar{h}$ in the order $\overline{T}\,\overline{M}M^*$ and let N^* denote a

point on $\bar{k}$ in the order $\bar{K}\bar{N}N^*$. Let t denote an arc that has M^* and N^* as endpoints, lies except for its endpoints entirely in $\bar{D}$ and has no point in common with $\bar{M}\bar{N}$. If a curve of G contains a point P of $\bar{M}\bar{N}$ there is an interval of that curve that contains P and has its endpoints on t. It follows that there exists an infinite set of distinct arcs $A_1 P_1 B_1$, $A_2 P_2 B_2$, $A_3 P_3 B_3$, $\cdots$ such that (1) for every n, A_n and B_n are on t and the arc $A_n P_n B_n$ is an interval of a curve of class III, (2) if n_1 is distinct from n_2, $A_{n_1} P_{n_1} B_{n_1}$ and $A_{n_2} P_{n_2} B_{n_2}$

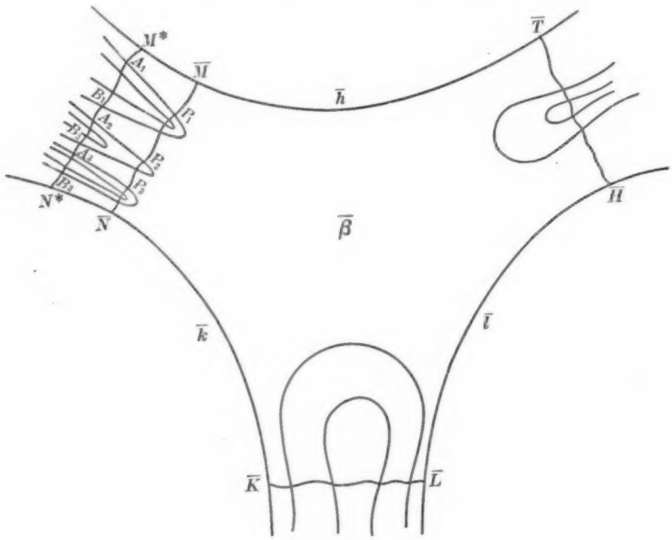

Fig. 6.

are not intervals of the same curve of the set G, (3) for every n the arc $A_n P_n B_n$ contains a point of $\bar{M}\bar{N}$, (4) if n_1 is distinct from n_2 the intervals $A_{n_1} B_{n_1}$ and $A_{n_2} B_{n_2}$ of $M^* N^*$ have no point in common. If ϵ is the least distance from a point of t to a point of $\bar{M}\bar{N}$ there exists a positive integer $\bar{n}$ such that the distance from $A_{\bar{n}}$ to $B_{\bar{n}}$ is less than $\delta_{t,\epsilon}$. But there exists an interval $A_{\bar{n}} P_{\bar{n}} B_{\bar{n}}$ of a curve of the set G with $A_{\bar{n}}$ and $B_{\bar{n}}$ as endpoints and containing a point $P_{\bar{n}}$ at a distance of ϵ or more from $A_{\bar{n}}$. Thus the supposition that no one of the curves h, k and l separates the other two has led to a contradiction. It follows that, of any three curves of G, one separates the other two.

If $\bar{g}$ is a definite G-curve, C is a definite circle and A is a point of C not lying on $\bar{g}$ there exists a G-curve which lies on the A-side of $\bar{g}$ but contains

* See Theorem 2 of my paper *On the most general class L of Fréchet in which the Heine-Borel-Lebesgue Theorem holds true*, Proceedings of the National Academy of Sciences, vol. 5 (1919), p. 208.

no point of C. This may be proved as follows. For every G-curve g that lies on the A-side of $\bar{g}$ and contains a point of C let C_g denote the set of all points $[X]$ of C such that X is either on g or on the far side of g from $\bar{g}$. For every such g the set C_g is closed and bounded and for every two such g's, $\bar{g}_1$ and $\bar{g}_2$, either $C_{\bar{g}_1}$ contains $C_{\bar{g}_2}$ or $C_{\bar{g}_2}$ contains $C_{\bar{g}_1}$. It follows* that there exists at least one point P which belongs to C_g for every G-curve g which lies on the A-side of $\bar{g}$ and contains a point of C. If g_P denotes that G-curve which contains the point P then no G-curve which lies on the far side of g_P from $\bar{g}$ can contain any point of C.

It easily follows that for every circle C there exist two G-curves such that every point of C lies between them. Now let O denote some definite point and for each positive integer n let C_n denote a circle with center at O and radius n. Let g_0 denote that G-curve which passes through O. Let g_1 and g_{-1} denote two G-curves such that C_1 lies between them. Let g_2 and g_{-2} denote two G-curves such that C_2 lies between them and such that g_2 is on the far side of g_1 from O and g_{-2} is on the far side of g_{-1} from O. This process may be continued. It follows that there exists a set G_0 of G-curves consisting of two infinite sequences $g_0, g_1, g_2, \cdots$ and $g_{-1}, g_{-2}, g_{-3}, \cdots$ such that, for each positive n, C_n lies between g_n and g_{-n}, g_{n+1} is on the far side of g_n from O and $g_{-(n+1)}$ is on the far side of g_{-n} from O. It is clear that every point is either on some curve of the set G_0 or between two successive curves of G_0. With the use of the fact that the system G is inversely equicontinuous with respect to every bounded point-set and that, of any three curves of G, one separates the other two, it can be shown that there exist four infinite sequences of points $A_0, A_1, A_2, A_3, \cdots; A_{-1}, A_{-2}, A_{-3}, \cdots; B_0, B_1, B_2, B_3, \cdots$ and $B_{-1}, B_{-2}, B_{-3}, \cdots$ and two sequences of arcs $A_0 B_0$, $A_1 B_1$, $A_2 B_2$, $\cdots$ and $A_{-1} B_{-1}$, $A_{-2} B_{-2}$, $A_{-3} B_{-3}$, $\cdots$ such that (1) for every n the points A_n, A_{n+1}, A_{n+2} are in the order $A_n A_{n+1} A_{n+2}$ on g_1 and the points B_n, B_{n+1}, B_{n+2} are in the order $B_n B_{n+1} B_{n+2}$ on g_0, (2) for each n and $m\,(m \neq n)$ the arcs $A_n B_n$ and $A_m B_m$ lie entirely between g_0 and g_1 and have no point in common, (3) for every point X on g_1 and every point Y on g_0 there exists a positive integer n such that X is on the interval $A_{-n} A_n$ of g_1 and Y is on the interval $B_{-n} B_n$ of g_0, (4) if, for each n, J_n denotes the closed curve formed by the arcs $A_n B_n$, $A_{n+1} B_{n+1}$ and the G-intervals $A_n A_{n+1}$ and $B_n B_{n+1}$ then (a) every point between g_0 and g_1 is on or within some J_n, (b) if $|m - n| > 1$ every G-interval whose endpoints are on or within J_m lies wholly without J_n. For each integer n let K_n denote the set of all points $[X]$ such that X lies on a G-interval whose endpoints are within J_{4n}. By methods wholly or largely identical with those employed in the proof of Lemma 1 it may be shown that there exists an arc $D_n E_n$ which lies entirely in the domain K_n, except that its endpoints D_n and E_n lie on g_1 and g_0 respectively,

and which does not have more than one point in common with any curve of the set G. For each n let $\bar{J}_n$ denote the closed curve found by the arcs $D_n E_n$, $D_{n+1} E_{n+1}$ and the G-intervals $D_n D_{n+1}$, $E_n E_{n+1}$ and let $\bar{R}_n$ denote its interior. By Theorem 1 there exists a set of arcs α_n such that (1) each arc of α_n has its endpoints on g_1 and g_0 respectively and lies, except for its endpoints, wholly within $\bar{J}_n$, (2) no two arcs of α_n have a point in common, (3) through each point of the point-set composed of $\bar{R}_n$ and the two G-segments $D_n D_{n+1}$ and $E_n E_{n+1}$ there is one and only one arc of the set α_n, (4) no arc of α_n has more than one point in common with any one arc of the set G. Let H_0 denote the set of arcs composed of all the arcs of all the sets α_n together with all the arcs $D_n E_n$. For each n there exists a set of arcs H_n bearing to g_n and g_{n+1} a relation similar to the above described relation of H_0 to g_0 and g_1, so that (1) each arc of H_n lies entirely between g_n and g_{n+1} except that its endpoints are on g_n and g_{n+1} respectively, (2) through each point that lies on g_n or g_{n+1} or between them there is just one arc of H_n, (3) no arc of H_n has more than one point in common with any arc of the set G. For each point P there exists n_P such that P is either on g_{n_P} or between g_{n_P} and g_{n_P+1}. Let h_{1P} denote that arc of H_{n_P} which passes through P. Let h_{2P} denote that arc of H_{n_P+1} which has an endpoint in common with h_{1P} and let h_{0P} denote that arc of H_{n_P-1} which has an endpoint in common with h_{1P}. This process may be continued. Thus there exists a set of arcs $[h_{mP}]$ $(-\infty < m < \infty)$ such that, for every m, $h_{(m+1)P}$ belongs to the set H_{n_P+m} and has an endpoint in common with $h_{(m+2)P}$. The point-set obtained by adding together all the arcs of the set $[h_{mP}]$ is an open curve h_P that passes through the point P and has just one point in common with each curve of the set G. Let H denote the set of all curves h_P for all points P of S. Through each point of S there is just one curve of the set H and just one curve of the set G and if h is any curve of H and g is any curve of G, h and g have just one point in common. It follows* that there exists a one to one transformation of S into itself which carries H into a complete system of parallel lines and G into another complete system of parallel lines.

THEOREM 3. *If $A\bar{A}_0 \bar{B}_0 B$ is a rectangle and $A_1 B_1$, $A_2 B_2$, $A_3 B_3$, $\cdots$ is an infinite sequence G of arcs such that* (1) *the points A_1, A_2, A_3, $\cdots$ are in the order $\bar{A}_0 A_1 A_2 A_3 \cdots A_n A_{n+1} \cdots A$ on the interval $\bar{A}_0 A$ and the points B_1, B_2, B_3, $\cdots$ are in the order $\bar{B}_0 B_1 B_2 B_3 \cdots B_n B_{n+1} \cdots B$ on the interval $\bar{B}_0 B$,* (2) *every arc of G lies except for its endpoints entirely within the rectangle $A\bar{A}_0 \bar{B}_0 B$,* (3) *no two arcs of G have a point in common and* (4) *for each positive number ϵ there exists a positive number n_ϵ such that if $n > n_\epsilon$, then every point of $A_n B_n$ is at a distance less than ϵ from the line AB; then in order that the sequence G*

* Cf. pp. 177–178 of my paper *Concerning a set of postulates for plane analysis situs*, loc. cit.

should be equivalent from the standpoint of analysis situs to an infinite sequence of straight line intervals, satisfying the same conditions (1)–(4), *and all parallel to* AB, *it is necessary and sufficient that the set of arcs* G *should be equicontinuous.*

That this condition is necessary, is evident. I will show that it is sufficient.

Proof. Suppose G is an equicontinuous sequence of arcs satisfying conditions (1)–(4) of the hypothesis of Theorem 3. By hypothesis for every positive number ϵ there exists a positive number δ_ϵ such that if P_1 and P_2 are two points on an arc g of G at a distance apart less than or equal to δ_ϵ then the interval $P_1 P_2$ of g lies entirely within some circle of radius ϵ. It follows with the help of condition (4) that if X and Y are two points of AB at a distance apart less than or equal to δ_ϵ and p_1 and p_2 are straight lines perpendicular to AB at X and Y respectively then if* $n > n_{\delta_\epsilon}$ no interval of $A_n B_n$ with endpoints on p_1 contains a point of p_2. If, for every n, X_n denotes the last point that $A_n B_n$ has in common with p_1 and Y_n denotes the first point that it has in common with p_2 it follows that if n_1 and n_2 are positive integers greater than n_{δ_ϵ} and P_{n_1} and P_{n_2} are points between p_1 and p_2 on the intervals $X_{n_1} Y_{n_1}$ and $X_{n_2} Y_{n_2}$ respectively of the arcs $A_{n_1} B_{n_1}$, $A_{n_2} B_{n_2}$ then P_{n_1} can be joined to P_{n_2} by a simple continuous arc that lies wholly between p_1 and p_2 and lies except for its endpoints wholly between the arcs $A_{n_1} B_{n_1}$ and $A_{n_2} B_{n_2}$. Now for each positive integer n subdivide the interval AB into 3^n equal subintervals by $3^n - 1$ points $A_{n1}, A_{n2}, A_{n3}, \cdots, A_{n(3^n-1)}$ $(1 \leqq n < \infty)$ in the order $AA_{n_1} A_{n_2} \cdots A_{n(3^n-1)} B$. For each n and m $(1 \leqq m \leqq 3^n - 1)$ let p_{nm} denote the perpendicular to AB at the point A_{nm}. There exists a sequence of positive integers $\bar{n}_1, \bar{n}_2, \bar{n}_3, \cdots$ such that $\bar{n}_1 < \bar{n}_2 < \bar{n}_3 \cdots$ and such that, for every k, $\bar{n}_k > n_{\delta_{l/3^k}}$, where l is the length of AB. For each k let $\bar{g}_k$ denote the arc $A_{\bar{n}_k} B_{\bar{n}_k}$ and let $\bar{A}_k$ and $\bar{B}_k$ denote its endpoints, $\bar{A}_k$ being that one which lies on $A\bar{A}_0$. For each n and m $(1 \leqq m \leqq 3^n - 1)$ let B_{nm} be the first and A_{nm} the last point that the arc $\bar{g}_n$ has in common with p_{nm}. Let t_{nm} denote the G-segment $A_{nm} B_{n(m+1)}$. For each n and each positive integer m (less than 3^n) of the form $3k - 2$ (where k is an integer) let X_{nm} denote a point of the segment t_{nm}. If, for each such n and m, $\overline{m}$ denotes the number $3m + 1$, there exists (Fig. 7) an arc $X_{nm} X_{(n+1)\overline{m}}$ which has not† more than one point in common with any arc of the set G, lies wholly between the lines p_{nm} and $p_{n(m+1)}$ and also lies, except for its endpoints, wholly between the arcs $\bar{g}_n$ and $\bar{g}_{n+1}$. For every n $(0 \leqq n < \infty)$ let A'_n denote a point on the straight line interval $\bar{A}_0 A$ at a distance from A equal to $a/(n+1)$, where a is the length of $\bar{A}_0 A$, and let B'_n denote a point on the interval $\bar{B}_0 B$ at the distance $a/(n+1)$ from B. Let $\bar{g}_0$ denote the straight line interval $\bar{A}_0 \bar{B}_0$. For each n $(0 \leqq n < \infty)$ let J_n denote the closed curve formed by the arcs $\bar{g}_n$ and $\bar{g}_{n+1}$ and the intervals

* For the meaning of n_{δ_ϵ} see Condition (4).

† There are not more than a finite number of arcs of the set G between $\bar{g}_n$ and $\bar{g}_{n+1}$.

$\overline{A}_n \overline{A}_{n+1}$ and $\overline{B}_n \overline{B}_{n+1}$ of $A\overline{A}_0$ and $B\overline{B}_0$ respectively. Let R_n denote the point-set composed of J_n and its interior. Let R'_n denote the point-set composed of the rectangle $A'_n A'_{n+1} B'_{n+1} B'_n$ and its interior and let R be that composed of $A\overline{A}_0 \overline{B}_0 B$ and its interior. Let b denote the length of AB. With the aid of a theorem of Schoenflies'* it may be easily seen that there exists a sequence of one to one transformations $T_0, T_1, T_2, \cdots$ such that, for each n ($0 \leqq n < \infty$), (1) T_n is a continuous transformation of R_n into R'_n, (2) T_n trans-

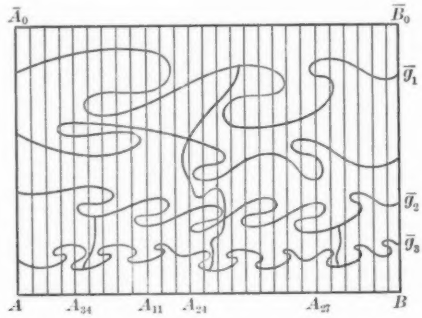

FIG. 7.

forms $\overline{A}_n$, $\overline{A}_{n+1}$, $\overline{B}_n$ and $\overline{B}_{n+1}$ into A'_n, A'_{n+1}, B'_n and B'_{n+1} respectively, (3) if P is a point of $\overline{g}_{n+1}$, $T_n(P) = T_{n+1}(P)$, (4) if there is any G-arc between $\overline{g}_n$ and $\overline{g}_{n+1}$ every such arc is transformed by T_n into a straight line interval parallel to AB, (5) if $n \geqq 1$ then for each m (less than 3^n) of the form $3k - 2$, where k is a positive integer, the point X_{nm} is transformed by T_n into a point X'_{nm} lying on the straight interval $A'_n B'_n$ at a distance from $A\overline{A}_0$ equal to $(m + 1/2)b/3^n$ and the arc $X_{nm} X_{(n+1)\overline{m}}$ is transformed into the straight line interval joining the point X'_{nm} to the point $X'_{(n+1)\overline{m}}$. For each n let H_n denote the set of all arcs $[h]$ in R_n such that $T_n(h)$ is a vertical† straight line interval. If $\overline{P}$ is a point on $\overline{A}_0 \overline{B}_0$ let $h_{\overline{P}0}$ denote that arc of H_0 which contains $\overline{P}$, let $h_{\overline{P}1}$ denote that arc of H_1 which has an endpoint in common with $h_{\overline{P}0}$, let $h_{\overline{P}2}$ denote that arc of H_2 which has an endpoint in common with $h_{\overline{P}1}$, and so on indefinitely. It is possible to show that there exists only one point $O_{\overline{P}}$ on AB which is a limit point of the point-set $h_{\overline{P}0} + h_{\overline{P}1} + h_{\overline{P}2} + \cdots$ and that the set of points $O_{\overline{P}} + h_{\overline{P}0} + h_{\overline{P}1} + h_{\overline{P}2} + \cdots$ is a simple continuous arc from $\overline{P}$ to $O_{\overline{P}}$. Let H denote the set of all such arcs for all points $\overline{P}$ on $\overline{A}_0 \overline{B}_0$. Let K denote the set of arcs composed of AB and every arc in R which, for some n, is transformed by T_n into a straight interval parallel to AB. If P is any point of R let O_P denote the point which AB has in common with

* *Bericht über die Entwickelung der Lehre von den Punktmannigfaltigkeiten*, Part II, p. 108.

† I.e., perpendicular to AB.

that arc of H which passes through P and let L_P denote the point which $\overline{A_0 A}$ has in common with that arc of K which passes through P. For each point P, of R, let $T(P)$ denote the point in which the perpendicular to AB at the point O_P intersects the perpendicular to $\overline{AA_0}$ at the point L_P. The so determined transformation T is a continuous transformation of R into itself. It is easy to see that there exists a continuous transformation of S into itself which reduces to T on R. Every such transformation satisfies the requirements of Theorem 3.

The truth of the following theorems may also be established.

THEOREM 4. *If, in a plane S, O is a point and G is a set of open curves through O such that through each point of S distinct from O there is one and only one curve of the set G, then in order that G should be equivalent from the standpoint of analysis situs to the set of all straight lines in S through O it is necessary and sufficient that G should be equicontinuous with respect to every bounded set of points.*

THEOREM 5. *If, in a plane S, O is a point and G is a set of simple closed curves enclosing O such that through each point of S distinct from O there is one and only one curve of the set G, then in order that G should be equivalent from the standpoint of analysis situs to the set of all circles in S with center at O it is necessary and sufficient that the set G should be equicontinuous with respect to every bounded set of points.*

UNIVERSITY OF TEXAS

FUNDAMENTAL SYSTEMS OF FORMAL MODULAR SEMINVARIANTS OF THE BINARY CUBIC*

BY

W. L. G. WILLIAMS

Introduction

The present paper owes its origin to an attempt to solve the problem of Hurwitz in the case of the binary cubic form, viz., to find a set of formal modular invariants of the binary cubic form such that all others could be expressed as polynomials with integral coefficients with these as arguments. In this attempt I have not yet been successful, but methods have been developed which will aid in the solution of the problem and which have resulted in the solution of the similar, though less difficult problem of the determination of a fundamental system of formal modular seminvariants of the binary cubic form, modulis 5 and 7. These methods are so general in their nature that they could easily be used in the determination of a like system with respect to any prime modulus greater than three.

The problem of a fundamental system of the seminvariants of the cubic here solved in the case of the moduli 5 and 7 has already been solved by Dickson in the case of the modulus 5 and the methods which he used in that case could no doubt have been used for its solution in other cases. Whatever interest may attach to the present paper does not then lie in the fact that a solution exists nor yet in the fundamental system exhibited, but in whatever power or generality the methods used in the solution may have, and in the fact, hitherto unknown, that in the case of the cubic the number of members of a fundamental system is a function of the modulus instead of being a constant as in the case of the quadratic.

The method of annihilators, fundamental in the classical theory of invariantive concomitants, seems at first to be almost useless in the formal modular case. This is due to the fact that not only are these annihilators not linear but they are of an order which depends upon the modulus. Their complexity does indeed make the computation of seminvariants and invariants through their use very laborious in any except the simplest cases, but they furnish simple proofs of general theorems of far reaching significance.

* Presented to the Society, October 30, 1920.

A second method of importance is founded on the use of sums and products of linear functions of the coefficients of the form. This method is peculiarly adaptable to the modular in contrast to the algebraic case. Although this method has been previously used by Dickson and Hurwitz for the representation of certain invariants and seminvariants, in the present paper the interesting fact appears that all non-algebraic seminvariants of the cubic for the moduli considered can be easily and elegantly represented in this way, and one is strongly tempted to generalize and to say that this is true of all formal modular seminvariants, and to call this symmetric form their "natural" form.

In the algebraic theory many interesting developments have arisen from the arrangement of seminvariants and invariants in descending powers of the most advanced coefficient. An analogous method in the case of covariants has been used with marked success in the papers of Glenn on the covariants of binary forms. It is by the use of this method of leaders, now used so far as I know for the first time in the case of formal modular seminvariants and invariants, that the completeness of the system exhibited is proved. This part of the paper will no doubt appear to the reader as to the author tedious and awkward. The method of leaders is useful and elegant so long as it is applied in a search for concomitants and in their classification, but there is great need of more direct and powerful methods in the proofs of the completeness of systems.

In Section D, which is in the nature of a postscript to the paper, a fundamental system of protomorphic formal modular seminvariants is derived. In the algebraic case the fundamental system of seminvariants has only four members, in the formal modular case, modulo 5, this number is increased to 12, and when the modulus is 7, to 20; in the case of higher moduli the number is enormous. The protomorphs present a strong contrast; in the algebraic case the number of protomorphs in a fundamental system is 3, and by the addition to these three of the single seminvariant

$$\beta = \prod_{t=0}^{p-1} (at + b) \equiv a^{p-1} b - b^p \ (\text{modulo } p)$$

we obtain a fundamental system of protomorphs for any prime greater than three.

A. General theorems

If a binary form

$$f(x, y) = a_0 x^n + n a_1 x^{n-1} y + \cdots + a_n y^n \quad (n \not\equiv 0, \text{modulo } p),$$

in which $a_0, a_1, \cdots, a_n$ are arbitrary variables, be transformed by the substitution

$$(1) \qquad \begin{aligned} x &= lX + mY, \\ y &= l' X + m' Y \end{aligned}$$

(l, m, l', m' being integers, taken modulo p) of determinant

$$D = \begin{vmatrix} l & m \\ l' & m' \end{vmatrix} \not\equiv 0 \text{ (modulo } p\text{)},$$

a binary n-ic form

$$A_0 X^n + n A_1 X^{n-1} Y + \cdots + A_n Y^n$$

results, in which

$$A_0 = f(l, l'),$$
$$A_1 = l^{n-1} m a_0 + \cdots,$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot,$$
$$A_n = f(m, m').$$

A polynomial $P(a_0, \cdots, a_n)$ for which

$$P(A_0, \cdots, A_n) \equiv D^\lambda P(a_0, a_1, \cdots, a_n) \text{ (modulo } p\text{)}$$

identically as to $a_0, \cdots, a_n$ after the A's have been replaced by their values in terms of the a's, under all transformations (1) is called a formal modular invariant, modulo p, of f. In like manner a polynomial

$$Q(a_0, a_1, a_2, \cdots, a_n)$$

which is unchanged, modulo p, by the transformation induced by all substitutions (in which t is integral)

$$x = X + tY,$$
$$y = \quad\quad Y$$

of determinant unity is called a formal modular seminvariant of $f(x, y)$ modulo p. It is clear that all algebraic invariants and seminvariants are also formal modular invariants and seminvariants. In this paper formal modular seminvariants and invariants will be frequently referred to as formal seminvariants and invariants, or simply as seminvariants and invariants.

Invariants of this type were first considered by Hurwitz.* L. E. Dickson† in his *Madison Colloquium Lectures* first exhibited a fundamental system‡ of formal invariants and seminvariants of the binary quadratic form, modulo p, and a fundamental system of seminvariants of the binary cubic, modulo 5.

The object of the present paper is to derive a fundamental system of seminvariants for the same cubic form and the same modulus 5 (practically identical with the system of Dickson) by a different method and to apply the method

* Archiv der Mathematik und Physik (3), vol. 5 (1903).

† *The Madison Colloquium Lectures on Mathematics*, pp. 40–53.

‡ By a fundamental system is meant, as in the algebraic theory, a set of invariants (seminvariants) $S_0, \cdots, S_r$ such that any invariant (seminvariant) can be expressed as a polynomial $P(S_0, \cdots, S_r)$.

to the case of the modulus 7. The present method is general and in the case of several of the seminvariants a general form will be given.

I

In the theory of formal modular concomitants weight is defined exactly as in the algebraic case and plays the same prominent part. In the algebraic theory a concomitant all of whose terms are of the same weight is said to be isobaric and all concomitants which are not homogeneous and isobaric can be expressed as the sums of concomitants which have these properties. In like manner in the present theory a concomitant the weights of whose terms are congruent to each other with respect to the modulus q is said to be modularly isobaric, modulo q; concomitants of a binary form modulo p which are not homogeneous and modularly isobaric with respect to the modulus $p-1$ can be expressed as sums of concomitants which have these properties. In this paper we confine our attention to such seminvariants and invariants.

II. An annihilator of formal seminvariants

Theorem. *A necessary and sufficient condition that $F(a_0, \cdots, a_n)$, a polynomial with integral coefficients, should be a formal seminvariant of the binary form*

$$a_0 x^n + na_1 x^{n-1} y + \cdots + a_n y^n$$

is that $\Theta F \equiv 0$ modulo p, where Θ is the operator

$$\Theta = \Omega + \frac{\Omega^p}{p!} + \frac{\Omega^{2p-1}}{(2p-1)!} + \cdots,$$

where

$$\Omega = a_0 \frac{\partial}{\partial a_1} + 2a_1 \frac{\partial}{\partial a_2} + \cdots + na_{n-1} \frac{\partial}{\partial a_n}$$

and Ω^q is its qth iteration, while $1, p, 2p-1, \cdots$ have the common difference $p-1$.

Proof. If we apply to the given form the substitution]

$$x = X + tY, \qquad y = Y \qquad (t \not\equiv 0, \text{ modulo } p)$$

we derive the equivalent substitutions

$$\begin{aligned} A_0 &= a_0, \\ A_1 &= a_1 + a_0 t, \\ A_2 &= a_2 + 2a_1 t + a_0 t^2, \\ &\cdots\cdots\cdots\cdots, \\ A_n &= a_n + na_{n-1} t + \frac{n(n-1)}{2!} a_{n-2} t^2 + \cdots + a_0 t^n, \end{aligned}$$

where $A_0, A_1, A_2, \cdots, A_n$ are the coefficients in the transformed quantic. Expanding $F(A_0, A_1, \cdots, A_n)$ by Taylor's theorem* in powers of t and reducing the powers of t higher than the $(p-1)$th by Fermat's theorem (modulo p) we have

$$F(A_0, A_1, \cdots, A_n) \equiv F(a_0, a_1, \cdots, a_n)$$
$$+ t\left(\Omega + \frac{\Omega^p}{p!} + \cdots\right) + \cdots + t^{p-1}\left(\frac{\Omega^{p-1}}{(p-1)!} + \cdots\right) \quad (\text{modulo } p).$$

Now a necessary and sufficient condition that $F(A_0, A_1, \cdots, A_n)$ be independent of t and so $\equiv F(a_0, a_1, \cdots, a_n)$, modulo p, which it is when $t \equiv 0$, modulo p, is that $\partial F/\partial t \equiv 0$, modulo p, and we see that $\partial F/\partial t = \Theta F$, whence the theorem follows.

Remark. It will be evident that $\Omega^p F = p!\Phi$ where Φ is a polynomial in $a_0, a_1, \cdots, a_n$ with integral coefficients. By the symbol $(\Omega^p/p!)F$ we mean the polynomial Φ, it being understood that the division by $p!$ has been performed. Similar remarks apply to the other terms of Θ. In the use of this annihilator of formal modular seminvariants no reduction with respect to the modulus is allowable until ΘF has been calculated without such reduction. Practice in the use of this annihilator will aid in the understanding of the proofs which follow and for this reason an example is added.

Example. To show that

$$K \equiv -b^4 - a^2bd + a^2c^2 + ab^2c - c^4 + bc^2d + b^2d^2 - acd^2$$

is a seminvariant, modulo 5, of the binary cubic,

$$ax^3 + 3bx^2y + 3cxy^2 + dy^3.$$

In this case

$$\Omega = a\frac{\partial}{\partial b} + 2b\frac{\partial}{\partial c} + 3c\frac{\partial}{\partial d}$$

and

$$\Theta = \Omega + \frac{\Omega^5}{5!} + \frac{\Omega^9}{9!} + \cdots;$$

however since $\Omega^9 K, \Omega^{13} K, \cdots = 0$,

$$\Theta K = \Omega K + \frac{\Omega^5 K}{5!}.$$

* Cf.: E. B. Elliott: *Algebra of Quantics*, p. 114. L. E. Dickson: These Transactions, vol. 8 (1907), p. 209. O. E. Glenn: American Journal of Mathematics, vol. 37 (1915), p. 73.

$$\Omega K = -2ab^3 + 3a^2bc - a^3d - 5ac^2d + 10b^2cd - 5bc^3,$$

$$\frac{\Omega^2 K}{20} = -ac^3 + b^3 d,$$

$$\frac{\Omega^3 K}{20} = -6abc^2 + 3ab^2 d + 3b^3 c,$$

$$\frac{\Omega^4 K}{5!} = -ab^2 c + a^2 bd + b^4 - a^2 c^2,$$

$$\frac{\Omega^5 K}{5!} = -3a^2 bc + 2ab^3 + a^3 d,$$

whence $\Theta K \equiv 0$, mod 5.

III

Definition. If a seminvariant S of the cubic

$$ax^3 + 3bx^2 y + 3cxy^2 + dy^3$$

be arranged in descending powers of d thus,

$$S = S_0 d^n + S_1 d^{n-1} + \cdots + S_n,$$

$S_0 d^n$ is called the leading term of the seminvariant and S_0 is called the leader of the seminvariant.

THEOREM. *The leader of any seminvariant of the cubic is a seminvariant of the quadratic when $n \not\equiv 0$, modulo p and either a seminvariant of the quadratic or a constant when $n \equiv 0$, modulo p.*

Proof.

$$\Theta S = (\Theta S_0) d^n + \left\{ \Theta S_1 + 3cnS_0 + \frac{3cn}{(p-1)!}(\Omega^{p-1} S_0) + \frac{6nb}{2!(p-2)!}(\Omega^{p-2} S_0) + \frac{6na}{3!(p-3)!}(\Omega^{p-3} S_0) + \cdots \right\} d^{n-1} + \cdots.$$

Since ΘS must vanish identically, modulo p,

$$\Theta S_0 \equiv 0 \quad (\text{modulo } p),$$

$$\Theta S_1 \equiv -3cnS_0 - \cdots \quad (\text{modulo } p),$$

which shows that S_0 is a constant or a seminvariant of the quadratic $ax^2 + 2bxy + cy^2$ [for it contains no terms involving d and is therefore annihilated by Θ when $\Omega = a(\partial/\partial b) + 2b(\partial/\partial c)$]. If $n \not\equiv 0$, modulo p, S_0 cannot be a constant for in that case ΘS would involve a constant multiple of cd^{n-1} and consequently ΘS would not be congruent to zero, modulo p.

If S_0 is an algebraic seminvariant and hence annihilated by Ω the second congruence above reduces to $\Theta S_1 \equiv 0$ (modulo p) when $n \equiv 0$, modulo p, and in this case S_1 is a seminvariant, considerations of homogeneity showing us that it cannot be a constant.

IV

If the cubic

$$F = ax^3 + 3bx^2 y + 3cxy^2 + dy^3$$

be changed into

$$F' = AX^3 + 3BX^2 Y + 3CXY^2 + DY^3$$

by the substitution

$$x = X + Y, \qquad y = Y,$$

then

$$A = a, \qquad B = b + a, \qquad C = c + 2b + a, \qquad D = d + 3c + 3b + a;$$

and the substitution

$$A = -d, \qquad D = a, \qquad B = c, \qquad C = -b,$$

is induced on the coefficients of F by

$$x = Y, \qquad y = -X.$$

Any function of a, b, c, d which is invariant under the first of these substitutions is a seminvariant of F, modulo p,* for if

$$F(a, b, c, d) \equiv F(a, b + a, c + 2b + a, d + 3c + 3b + a) \text{ (modulo } p\text{)},$$

repeating the substitution $(t - 1)$ times we have

$$F(a, b, c, d) \equiv F(a, b + at, c + 2bt + at^2, d + 3ct$$
$$+ 3bt^2 + at^3) \text{ (modulo } p\text{)} \quad (t = 1, 2, 3, \cdots, p - 1).$$

But these congruences simply state the truth of the condition that $F(a, b, c, d)$ should be a seminvariant according to the definition given above. All functions of a, b, c, d which are invariant under both sets of transformations are invariants of F, modulo p.†

Dickson has given

$$F_t(a, b, c, d) = a(t^3 - 3kt - j) + 3b(t^2 - k) + 3ct + d$$
$$(j, k = 0, 1, 2, \cdots, p - 1)$$

* This fact is, of course, well known; I introduce a proof of it here because it is fundamental for all that follows and because I do not know where in the literature of the subject to refer the reader for a proof.

† *Madison Colloquium Lectures*, p. 49; these Transactions, vol. 8 (1907), pp. 207, 208.

as typical linear polynomials such that

$$F_t(a\,,\ b+a\,,\ c+2b+a\,,\ d+3c+3b+a) = F_{t+1}(a\,,b\,,c\,,d)\,.$$

It is easy to show that every linear polynomial having this property can be obtained from the given one by a proper choice of j and k, or is a constant multiple of one that can be so obtained. Dickson has also given

$$a(t^2-k)+2bt+c \qquad \text{and} \qquad at+b$$

as linear polynomials in $a\,,b\,,c$ and $a\,,b$, respectively, with similar properties and has pointed out that

$$\delta_{jk} = \prod_{t=0}^{p-1}\{a(t^3-3kt-j)+3b(t^2-k)+3ct+d\} \quad (j\,,k=0\,,\cdots,p-1)\,,$$

$$\gamma_k = \prod_{t=0}^{p-1}\{a(t^2-k)+2bt+c\} \qquad (k=0\,,\cdots,p-1)$$

and

$$\beta = \prod_{t=0}^{p-1}(at+b) \equiv b^p - a^{p-1}b \pmod p\,,$$

are seminvariants.* It is obvious that

$$\sum_{t=0}^{p-1} f[a(t^3-3kt-j)+3b(t^2-k)+3ct+d\,,\,a(t^2-k)+2bt+c\,,\,at+b]$$

are also seminvariants, f being any polynomial in its arguments with integral coefficients.

V

THEOREM. *The sum of the coefficients of the powers of t whose exponents are congruent to zero, modulo $p-1$, the exponent zero itself excepted, in the expansion in powers of t of any polynomial in one or more of the functions $(at^3+3bt^2+3ct+d)$, $(at^2+2bt+c)$, $(at+b)$ is a seminvariant of the cubic, modulo p.*

Proof. By Article IV

$$\sum_{t=0}^{p-1} P(at^3+3bt^2+3ct+d\,,\ at^2+2bt+c\,,\ at+b)$$

is a seminvariant of the cubic, modulo p. Now suppose P to be expanded so that

$$P = S_0 + S_1 t + \cdots + S_n t^n\,.$$

Then

$$\sum_{t=0}^{p-1} P = \sum_{t=0}^{p-1} S_0 + S_1 \sum_{t=0}^{p-1} t + \cdots + S_n \sum_{t=0}^{p-1} t^n\,.$$

Since

$$\sum_{t=0}^{p-1} t^r \equiv 0\,, \text{ modulo } p\,,$$

* *Madison Colloquium Lectures*, pp. 43 and 49.

when $r = 0$ and when $r \not\equiv 0$, modulo $p - 1$, and

$$\sum_{t=0}^{p-1} t^r \equiv -1, \text{ modulo } p,$$

when $r \neq 0$, but is congruent to zero, modulo $p - 1$,* therefore

$$\sum_{t=0}^{p-1} P = -(S_{p-1} + S_{2(p-1)} + \cdots)$$

which proves the theorem.

If we ascribe to t the weight 1 and to a, b, c, d the weights $0, 1, 2, 3$ respectively, P is absolutely isobaric, and consequently $S_{p-1}, S_{2(p-1)}, \cdots$ differ in weight by multiples of $p - 1$. Therefore

$$\sum_0^{p-1} P \qquad \text{and} \qquad (S_{p-1} + S_{2(p-1)} + \cdots)$$

are modularly isobaric, modulo $p - 1$.

VI

THEOREM. *If any formal modular seminvariant S be operated upon with the differential operators Ω, $F = a(\partial/\partial c) + 3b(\partial/\partial d)$ and $\partial/\partial d$, formal modular seminvariants result.*

Proof.† $S = \sum(a^r b^s c^t d^u)$. Since S is a seminvariant

$$\sum(a^r b^s c^t d^u) \equiv \sum\{a^r(b+a)^s(c+2b+a)^t(d+3c+3b+a)^u\} \quad (\text{modulo } p).$$

Operating on this congruence successively with Ω, F, and $\partial/\partial d$ we have respectively

$$\begin{aligned}
&\sum(sa^{r+1} b^{s-1} c^t d^u + 2ta^r b^{s+1} c^{t-1} d^u + 3ua^r b^s c^{t+1} d^{u-1})\\
&\equiv \sum\{sa^{r+1}(b+a)^{s-1}(c+2b+a)^t(d+3c+3b+a)^u\\
(1)\qquad &+ 2ta^r(b+a)^{s+1}(c+2b+a)^{t-1}(d+3c+3b+a)^u\\
&+ 3ua^r(b+a)^s(c+2b+a)^{t+1}(d+3c+3b+a)^{u-1}\} \quad (\text{modulo } p),
\end{aligned}$$

$$\begin{aligned}
&\sum\{ta^{r+1} b^s c^{t-1} d^u + 3ua^r b^{s+1} c^t d^{u-1}\}\\
(2)\qquad &\equiv \sum\{ta^{r+1}(b+a)^s(c+2b+a)^{t-1}(d+3c+3b+a)^u\\
&+ 3ua^r(b+a)^{s+1}(c+2b+a)^t(d+3c+3b+a)^{u-1}\} \quad (\text{modulo } p),
\end{aligned}$$

* Glenn, these Transactions, vol. 20 (1919), p. 156. Vandiver, Annals of Mathematics, ser. 2, vol. 18 (1916), p. 105.

† The proof as here given does not hold for terms in which one or more of $r, s, t, u \equiv 0$, modulo p; the reader will have no difficulty in extending the proof to such cases.

and

$$\sum u a^r b^s c^t d^{u-1} \equiv \sum \{ u a^r (b+a)^s (c+2b+a)^t (d+3c \\ + 3b + a)^{u-1} \} \text{ (modulo } p). \tag{3}$$

The congruences (1), (2), and (3) demonstrate the truth of the theorem.

VII

THEOREM. *In any seminvariant of the cubic (modulo p)*

$$S = S_0 d^{pq+r} + S_1 d^{pq+r-1} + \cdots + S_r d^{pq} + \cdots + S_{pq+r} \quad (p > r > 0,\ q \geqq 0),$$

$$S_1 d^{pq+r-1}, \cdots, S_r d^{pq}$$

all occur and the terms of highest weight in them are of the same absolute weight as the term or terms of highest weight in $S_0 d^{pq+r}$.

Proof. Let $H_0, H_1, \cdots, H_r$ be the terms of highest weight in $S_0, S_1, \cdots, S_r$ respectively.

$$\begin{aligned}\Theta S = (\Theta S_0) d^{pq+r} &+ \{3(pq+r)cH_0 + \Omega H_1 + \text{terms of lower weight}\} d^{pq+r} \\ &+ \{3(pq+r-1)cH_1 + \Omega H_2 + \text{terms of lower weight}\} d^{pq+r-1} \\ &+ \cdots\cdots\cdots\cdots\cdots\cdots \\ &+ \{3(pq+1)cH_{r-1} + \Omega H_r + \text{terms of lower weight}\} d^{pq} \\ &+ \cdots\cdots\cdots\cdots\cdots\cdots.\end{aligned}$$

Since

$$\Theta S \equiv 0 \text{ (modulo } p),$$

equating the coefficients of $d^{pq+r-1}, \cdots, d^{pq}$ to zero, modulo p, and paying attention to weights, we have:

weight of cH_0 = weight of ΩH_1, i.e., weight of $H_0 d^{pq+r}$
= weight of $H_1 d^{pq+r-1}$,

weight of cH_1 = weight of ΩH_2, i.e., weight of $H_1 d^{pq+r-1}$
= weight of $H_2 d^{pq+r-2}$,

. . . . ,

weight of cH_{r-1} = weight of ΩH_r, i.e., weight of $H_{r-1} d^{pq+1}$
= weight of $H_r d^{pq}$,

whence we see that weight of $H_0 d^{pq+r}$ = weight of $H_r d^{pq}$.

The existence of H_0 necessitates the existence of $H_1, H_2, \cdots, H_r$.

VIII

THEOREM. *The seminvariants whose leading terms are* βd^{q-1} $(q \geqq 1)$, *where* β *is the seminvariant*

$$\prod_{t=0}^{p-1} (at + b) \equiv b^p - a^{p-1} b \text{ (modulo } p),$$

are sums of seminvariants whose leading terms are numerical multiples of

$$a^2 \Delta^{(p-3)/2} d^q, \qquad \textit{where} \qquad \Delta = b^2 - ac.$$

Proof. In the identity

$$a(at^2 + 2bt + c)^{p-2}(at^3 + 3bt^2 + 3ct + d)^q$$

$$+ 2\left\{(b^2 - ac)t + \frac{bc - ad}{2}\right\}(at^2 + 2bt + c)^{p-2}(at^3 + 3bt^2 + 3ct + d)^{q-1}$$

$$= (at + b)(at^2 + 2bt + c)^{p-1}(at^3 + 3bt^2 + 3ct + d)^{q-1} \qquad (q \geqq 1)$$

the coefficients of like powers of d on the two sides of the equality sign are identical. Furthermore, the sum of the coefficients of t^{p-1}, $t^{2(p-1)}$, $\cdots$ in the expansion of the first quantity on the left-hand side of the equality sign is a seminvariant, by Article V above; also the sum of the coefficients of the same powers of t in the expansion of the quantity on the right side is a seminvariant, and the same must be true in the case of the second quantity on the left-hand side in virtue of the identity. Let us call the seminvariants arising from the first, second, and third of these quantities S_1, S_2, S_3, respectively; then $S_1 + S_2 \equiv S_3$, modulo p. The coefficient of d^q in S_3 is evidently zero; consequently the leading terms of S_1 and S_2 differ only in sign. We propose to prove in Lemma I that the leading term of S_1 is

$$-\tfrac{1}{2}a^2 \Delta^{(p-3)/2} d^q \qquad (q \geqq 1),$$

and it will then be clear that the leading term of S_2 is

$$\tfrac{1}{2}a^2 \Delta^{(p-3)/2} d^q \qquad (q \geqq 1);$$

in Lemma II it will be shown that the leading term of S_3 is

$$\beta d^{q-1} \qquad (q \geqq 1).$$

The proof of these two lemmas will complete the proof of the theorem.

LEMMA I. *The leading term of* S_1 *is*

$$-\tfrac{1}{2}a^2 \Delta^{(p-3)/2} d^q \qquad (q \geqq 1).$$

Proof. The coefficient of d^q, the highest power of d occurring in S_1, is simply the coefficient of t^{p-1} in $a(at^2 + 2bt + c)^{p-2}$, and this coefficient is

a times the coefficient of t^{p-1} in $(at^2 + 2bt + c)^{p-2}$. We must now calculate this latter coefficient. Ascribing to t, a, b, c the same weights as previously, $(at^2 + 2bt + c)^{p-2}$ is (absolutely) isobaric, weight $2p - 4$. The term in t^{p-1} is then $kA_{p-3}t^{p-1}$, k being a constant and A_{p-3} a homogeneous, (absolutely) isobaric function of a, b, c of weight $p - 3$. kA_{p-3} is then a homogeneous, (absolutely) isobaric seminvariant of the quadratic and must be a function of a and Δ only.* The only such function of a and Δ of degree $p - 2$ and weight $p - 3$ is $ka\Delta^{(p-3)/2}$. As the coefficient of ab^{p-3} in the coefficient of t^{p-1} is $(p - 2)2^{p-3}$,

$$k = (p-2)2^{p-3} \equiv -2\frac{2^{p-1}}{2^2} \equiv -\tfrac{1}{2} \pmod{p}.$$

The lemma is now proved.

LEMMA II. *The leading term of S_3 is βd^{q-1}.*

Proof. The coefficient of d^{q-1}, the highest power of d occurring in S_3, is the sum of the coefficients of t^{p-1} and t^{2p-2} in

$$(at + b)(at^2 + 2bt + c)^{p-1}.$$

We propose to show that the sum of these coefficients is β.

Consider the expansion

$$(at^2 + 2bt + c)^p = a^p t^{2p} + A_1 t^{2p-1} + \cdots + A_p t^p + \cdots + A_{2p-2}.$$

Differentiating and dividing by $2p$, we have

$$(at^2 + 2bt + c)^{p-1}(at + b) = a^p t^{2p-1} + \frac{2p-1}{2p}A_1 t^{2p-2} + \cdots + \tfrac{1}{2}A_p t^{p-1} + \cdots + \frac{A_{2p-1}}{2p}.$$

Let us first determine A_p; this can be done by differentiating the next to last identity p times, setting $t = 0$ and dividing by $p!$ To do this we proceed as follows:

$$(at^2 + 2bt + c)^p = a^p(t + \alpha)^p(t + \gamma)^p,$$

where

$$\alpha = \frac{b + \sqrt{b^2 - ac}}{a}, \qquad \gamma = \frac{b - \sqrt{b^2 - ac}}{a},$$

$$\frac{\partial^p}{\partial t^p}(at^2 + 2bt + c)^p = a^p\left\{\frac{\partial^p}{\partial t^p}[(t + \alpha)^p(t + \gamma)^p]\right\}.$$

Applying Leibniz's theorem

$$\frac{\partial^p}{\partial t^p}[(t + \alpha)^p(t + \gamma)^p] = p!(t + \alpha)^p + p^2(\qquad) + p!(t + \gamma)^p.$$

Setting $t = 0$ and multiplying by a^p we have

* Dickson, *Madison Colloquium Lectures*, p. 42 et seq.

$$\left\{\frac{\partial^p}{\partial t^p}(at^2 + 2bt + c)^p\right\}_{t=0} = p!(\alpha^p + \gamma^p)a^p + p^2 a^p(\quad).$$

But

$$\alpha^p + \gamma^p = 2\frac{b^p}{a^p} + p(\quad).$$

Therefore

$$(at^2 + 2bt + c)^p = p!\,2b^p + p^2(\quad).$$

Dividing by $p!$, $A_p \equiv 2b^p$, modulo p, and the coefficient of t^{p-1} in the second expansion above, namely $\frac{1}{2}A_p \equiv b^p$, modulo p. It is immediately evident that $A_1 = 2pa^{p-1}b$. Therefore the coefficient of t^{2p-2}, viz.,

$$\frac{2p-1}{2p}A_1 \equiv -a^{p-1}b, \text{ modulo } p.$$

Hence the lemma is proved.

IX. Seminvariants led by a

THEOREM. *Any seminvariant led by a is either a itself or has the same leading term as $a(\delta_{00})^r$, where $r \geqq 1$, and δ_{00} is the seminvariant obtained by setting $j = 0$, $k = 0$ in the δ_{jk} mentioned in Article IV.*

Proof. Let a seminvariant led by a be

$$S = ad^q + A_1 d^{q-1} + \cdots + A_q.$$

Operating on this with Θ we get a term $3qacd^{q-1}$, unless $q \equiv 0$, modulo p. Supposing for the moment that $q \not\equiv 0$, modulo p, we see that another term must occur in the result of the operation which is congruent to $-3qacd^{q-1}$, modulo p, in order that the seminvariant may be annihilated. Such a term could only come from $-3qbcd^{q-1}$. As this term must occur in the original seminvariant it too is operated on by Θ; operating on it we get a term $-6qb^2d^{q-1}$, which could not be made to disappear in the attempted annihilation as a corresponding term could not arise in the operation. Thus it follows that S is not a seminvariant when $q \not\equiv 0$, modulo p. A seminvariant $ad^q + \cdots$ exists for every value of q such that $q \equiv 0$, modulo p; when $q = 0$, the seminvariant is a itself and when $q = pr$ $(r \geqq 1)$, it is $a(\delta_{00})^r$.

X. Seminvariants led by γ^r

Let such a seminvariant be $\gamma_0^r d^q + \cdots$. In this seminvariant there is a term $c^{pr}d^q$; operating on this with Θ we get $(q \not\equiv 0, \text{ mod } p)$ $3qc^{pr+1}d^{q-1}$ and such a term could arise in the operation in no other way. Consequently the supposed seminvariant cannot exist as it cannot be annihilated by Θ. When $q \equiv 0$, mod p, a seminvariant $\gamma_0^r d^q$ exists for every such value of q, but every such seminvariant must have the same leading term as $\gamma_0^r(\delta_{00})^{q/p}$ $(q \geqq 0)$.

XI. Seminvariants led by $a\gamma_0^r$

Let such a seminvariant be $a\gamma_0^r d^q + \cdots$. A term $ac^{pr} d^q$ appears in the seminvariant. Operating with Θ we get ($q \not\equiv 0$, modulo p) $3qac^{pr+1} d^{q-1}$ and another term must appear in the operation which will cancel with this if S is a seminvariant. This term must come from a multiple of $bc^{pr+1} d^{q-1}$. There must be another term of this sort in order that the seminvariant be annihilated; this last could only come from a multiple of $b^2 c^{pr-1} d^q$, but the existence of such a term would contradict the original hypothesis that the coefficient of d^q in the seminvariant was divisible by a. Consequently, for values of q not congruent to zero, modulo p, there exist no seminvariants whose leading terms are $a\gamma_0^r d^q$; for every value of q such that $q \equiv 0$, modulo p, there exists such a seminvariant, viz., $a\gamma_0^r$ or $a\gamma_0^r(\delta_{00})^{pr}$, according as $q = 0$ or $q = pr'$ ($r' \geqq 1$). Accordingly any seminvariant $a\gamma_0^r d^q + \cdots$ involving d has as a leading term the same leading term as $a\gamma_0^r(\delta_{00})^{pr}$.

XII. Seminvariants whose leading terms are d^q

A seminvariant $d^q + \cdots$ must be annihilated if operated upon with Θ. If $q \not\equiv 0$, modulo p, a term $3qcd^{q-1}$, which is not congruent to zero, modulo p, appears in the result of the operation. As this term cannot be obained in any other way $d^q + \cdots$ cannot be annihilated and consequently is not a seminvariant.

If $q \equiv 0$, modulo p, there exists a seminvariant $d^q + \cdots$ for every value of q, viz., $(\delta_{00})^{q/p}$. Thus we have shown that from δ_{00} we can construct a seminvariant with the same leading term as any existing seminvariant whose leading term is d^q.

B. A fundamental system of formal seminvariants of the cubic modulo 5

XIII. Seminvariants led by a^2; seminvariants led by a^r ($r \geqq 3$)

Seminvariants led by a^2 are the algebraic seminvariants,

$$S_3 = a^2 d - 3abc + 2b^3,$$

$$D = a^2 d^2 - 6abcd + 4b^3 d + 4ac^3 - 3b^2 c^2,$$

and the formal modular seminvariants* (modulo 5):

* I have followed the notation of Dickson (*Madison Colloquium Lectures*, p. 52) because his σ_6 and mine have the same leading term, and the leading term of my σ_8 differs only in sign from that of his.

$$\sigma_5 \equiv -\sum_{t=0}^{4}(at^2+2bt+c)^2(at^3+3bt^2+3ct+d)^3$$

$$\equiv a^2d^3+(abc+b^3)d^2+(b^2c^2+2ac^3+a^2b^2+4ac^3)d+3a^4b$$

$$+2a^2bc^2+2ab^3c+3b^5+4bc^4, \quad \text{mod } 5,$$

$$\sigma_6 \equiv -\sum_{t=0}^{4}(at^2+2bt+c)^2(at^3+3bt^2+3ct+d)^4$$

$$\equiv a^2d^4+(3abc+3b^3)d^3+(2b^2c^2+3a^3c+4ac^3+2a^2b^2)d^2$$

$$+(2a^4b+3a^2bc^2+3ab^3c+2b^5+bc^4)d+c^6+4a^2b^4+4a^4c^2$$

$$+a^2c^4+b^4c^2+a^6, \quad \text{mod } 5.$$

Now multiplying these seminvariants whose leading terms are a^2d^2, a^2d^3, and a^2d^4 by $(\delta_{00})^r$, we obtain seminvariants whose leading terms are a^2d^{1+5r}, a^2d^{2+5r}, a^2d^{3+5r}, a^2d^{4+5r} $(r \geqq 1)$. We have thus constructed seminvariants whose leading terms are a^2d^q $(q \geqq 1)$.

By multiplying these seminvariants by the proper power of a we can construct a seminvariant whose leading term is a^td^q, where t is any integer greater than or equal to 2. We have thus shown that any seminvariant led by a^2 or any higher power of a has the same leading term as a seminvariant which can be constructed from a, S_3, D, σ_5, σ_6, and δ_{00}.

XIV. Seminvariants led by Δ^r $(r \geqq 1)$

Seminvariants whose leading terms are Δd and Δd^2 are

$$\sigma_3 \equiv \frac{1}{2}\sum_{t=0}^{4}(at^3+3bt^2+3ct+d)^3 \equiv (b^2-ac)d+2a^2b+3bc^2, \quad \text{mod } 5,$$

and the invariant*

$$K \equiv \sum_{t=0}^{4}(at^3+3bt^2+3ct+d)^4+a^4$$

$$\equiv (b^2-ac)d^2+(bc^2-a^2b)d-b^4-c^4+a^2c^2+ab^2c, \quad \text{mod } 5.$$

By multiplying Δ, σ_3, and K by $(\delta_{00})^r$ we obtain seminvariants whose leading terms are Δd^{5r}, Δd^{5r+1}, and Δd^{5r+2} $(r \geqq 1)$. That no seminvariants exist whose leading terms are Δd^{5r+3} and Δd^{5r+4} is evident from Article VII above.

$\Delta\sigma_3$, ΔK, $\sigma_3 K$, and K^2 are seminvariants whose leading terms are $\Delta^2 d$,

* K as here given differs only in sign from the K given by Dickson: *Madison Colloquium Lectures*, p. 51. Cf. also: Hurwitz, Archiv der Mathematik und Physik (3), vol. 5 (1903), p. 25; Dickson, these Transactions, vol. 8 (1907), p. 221; ibid., vol. 10 (1909), p. 154, footnote; Dickson, Bulletin of the American Mathematical Society, vol. 14 (1908), p. 316.

$\Delta^2 d^2$, $\Delta^2 d^3$, $\Delta^2 d^4$. Multiplying these and Δ^2 by $(\delta_{00})^r$ we have seminvariants whose leading terms are $\Delta^2 d^q$ $(q \geqq 1)$.

By multiplying these seminvariants by the proper power of Δ we have (together with those just given) seminvariants whose leading terms are $\Delta^r d^q$ where r is any integer greater than or equal to 2. We have thus shown that any existing seminvariant led by any power of Δ, has the same leading term as a seminvariant which can be constructed from Δ, σ_3, K, and δ_{00}.

XV. Seminvariants led by $a^r \Delta^s$ $(r, s \geqq 1)$

Seminvariants with leading terms $a\Delta d$, $a\Delta d^2$, $a\Delta d^3$, and $a\Delta d^4$ are* $a\sigma_3$, aK,

$$G_1 \equiv \frac{1}{2} \sum_{t=0}^{4} (at^2 + 2bt + c)^3 (at^3 + 3bt^2 + 3ct + d)^3$$

$$\equiv a\Delta d^3 + (a^3 b - abc^2) d^2 + (2ac^4 - a^5 - ab^4 - a^2 b^2 c) d - bc^5 + b^5 c + a^4 bc - 2ab^3 c^2 + 2a^2 bc^3, \quad \text{mod } 5,$$

$$\sigma_7 \equiv \frac{1}{2} \sum_{t=0}^{4} (at^2 + 2bt + c)^3 (at^3 + 3bt^2 + 3ct + d)^4$$

$$\equiv a\Delta d^4 + (2abc^2 + 3a^3 b) d^3 + (3a^5 + 4ac^4 + 3a^2 b^2 c + 3ab^4) d^2 + (bc^5 - b^5 c - a^4 bc + 2ab^3 c^2 + 3a^2bc^3) d + (3c^7 + 3a^2 c^5 + ab^2 c^4 + 2b^4 c^3 + 4a^5 b^2 + 3ab^6 + 2a^3 b^2 c^2 + a^4 c^3 - a^2 b^4 c), \quad \text{mod } 5.$$

From these by multiplying by powers of δ_{00} and then by $a^{r-1} \Delta^{s-1}$ $(r, s \geqq 2)$ we obtain seminvariants having the same leading terms as any seminvariant led by any power of a multiplied by any power of Δ.

XVI. Seminvariants led by $a^r \Delta^s \gamma_0^t \beta^u$ $(r, s, t \geqq 0;\ u \geqq 1)$

We have shown in Article VIII above how to express a seminvariant $\beta d^{q-1} + \cdots$ $(q > 0)$ as the sum of seminvariants whose leading terms are numerical multiples of $a^2 \Delta^{(p-3)/2} d^q$. For the modulus 5 we may verify by actual multiplication that

$$\beta \equiv 2(a^2 \sigma_3 - \Delta S_3), \quad \text{mod } 5,$$

$$\beta d + \cdots \equiv a^2 K - \Delta D, \quad \text{mod } 5,$$

$$\beta d^2 + \cdots \equiv \Delta\sigma_5 - aG_1, \quad \text{mod } 5,$$

$$\beta d^3 + \cdots \equiv DK - \Delta\sigma_6, \quad \text{mod } 5,$$

$$\beta d^4 + \cdots \equiv \sigma_5 K - a^2 \Delta\delta_{00}, \quad \text{mod } 5.$$

Multiplying these by $(\delta_{00})^r$ $(r \geqq 1)$, we obtain (together with those just

* $2g_1 = G + a\delta_{00} + a\Delta\sigma_3 + 3a^3 S_3$, modulo 5, G being the invariant G of *Madison Colloquium Lectures*, p. 50.

given) seminvariants whose leading terms are βd^q $(q \geqq 0)$. Multiplying these by $a^r \Delta^s \gamma_0^t \beta^{u-1}$ $(r, s, t \geqq 0;\ u \geqq 1)$, we obtain seminvariants whose leading terms are $a^r \Delta^s \gamma_0^t \beta^u d^q$ $(q \geqq 0)$. Thus we see that from a, Δ, S_3, σ_3, D, K, σ_5, σ_6, γ_0, and δ_{00} we can make up seminvariants having the same leading terms as any seminvariant led by $a^r \Delta^s \gamma_0^t \beta^u$.

XVII. Seminvariants led by $a^r \gamma_0^t$; seminvariants led by $\Delta^s \gamma_0^t$; seminvariants led by $a^r \Delta^s \gamma_0^t$

1. We have already treated the case of $a^r \gamma_0^t$ when $r = 1$ in XI above and we have shown in XIII how to form seminvariants whose leading terms are $a^r d^q$ $(r \geqq 2;\ q \geqq 1)$. Multiplying these by γ_0^t $(t \geqq 1)$ we have seminvariants with the same leading terms as all seminvariants led by $a^r \gamma_0^t$.

2. We have shown in Article XIV how to construct seminvariants whose leading terms are $\Delta d^{5r'}$, $\Delta d^{5r'+1}$, $\Delta d^{5r'+2}$, $\Delta^s d^q$ $(r', q \geqq 0;\ s \geqq 2)$. Multiplying these by γ_0^t we construct seminvariants whose leading terms are $\Delta^s \gamma_0^t d^{5r'}$, $\Delta^s \gamma_0^t d^{5r'+1}$, $\Delta^s \gamma_0^t d^{5r'+2}$, and $\Delta^s \gamma_0^t d^q$. It is easy to show by the use of Article VII that no seminvariants exist with leading terms $\Delta^s \gamma_0^t d^{5r'+3}$ and $\Delta^s \gamma_0^t d^{5r'+4}$. Thus from Δ, σ_3, K, γ_0, and δ_{00} we can make up seminvariants with the same leading terms as any existing seminvariants led by $\Delta^s \gamma_0^t$.

3. We have shown in Article XIV how to construct seminvariants with the same leading terms as any existing seminvariants led by $a^r \Delta^s$ $(r, s \geqq 1)$. Multiplying these by γ_0^t we obtain seminvariants with the same leading terms as any existing seminvariants led by $a^r \Delta^s \gamma_0^t$.

XVIII. Proof that the twelve seminvariants a, Δ, γ_0, S_3, D, σ_3, K, σ_5, σ_6, σ_7, G_1, and δ_{00} form a fundamental system of the cubic, modulo 5

We propose to prove this theorem by showing how actually to construct from these twelve seminvariants any rational integral homogeneous modularly isobaric (modulo 4) seminvariant of the cubic, modulo 5. It has been shown in Article III that any seminvariant S of the cubic is of the form

$$S = (S_0 + S_1 + \cdots + S_n) d^q + \cdots,$$

where S_0, S_1, $\cdots$, S_n are seminvariants of the quadratic (mod 5) or constants. It is this seminvariant S so arranged that we propose to construct. If one of S_0, S_1, $\cdots$, S_n is a constant, consideration of homogeneity shows that all the others are constants and that S has the same leader as a power of δ_{00}. Subtracting the proper numerical multiple of the proper power of δ_{00} from S we have a seminvariant which involves no higher power of d than d^{q-1}. If on the other hand S_0, S_1, $\cdots$, S_n be seminvariants of the quadratic,

each of them is a polynomial in a, Δ, β, and γ_0. We now consider this case, first when $q \equiv 0$, modulo 5, and second, when $q \not\equiv 0$, modulo 5.

(1) *When* $q \equiv 0$, *modulo* 5. The leader $S_0 + S_1 + \cdots + S_n$ of the seminvariant is a rational integral function of a, Δ, β, γ_0; constructing it from these and multiplying by $(\delta_{00})^q$ $(q = 5, 10, 15, \cdots)$, we construct from a, Δ, β, γ_0, and δ_{00} a seminvariant whose leading term is the same as that of S. Subtracting this from S we have a seminvariant which involves no power of d higher than d^{q-1}.

(2) *When* $q \not\equiv 0$, *modulo* 5. If q is equal to or greater than 1 none of the summands of the leader of S can be γ_0^t or $a\gamma_0^t$ for reasons similar to those given in Articles X and XI for the non-existence of seminvariants with these leaders $(t \geqq 1)$.

We have shown in Articles XVI and XVII how to construct from a, S_3, σ_3, D, K, σ_5, σ_6, G_1, and δ_{00} seminvariants whose leading terms are $a^r \Delta^s \gamma_0^t \beta^u d^q$ $(q \geqq 1; t \geqq 1; r, s, u$ ranging over all integral values except ones which will give the terms $a\gamma_0^t d^q$ and $\gamma_0^t d^q)$. Then subtracting from S seminvariants with the proper leaders made up from a, Δ, γ_0, S_3, σ_3, D, K, σ_5, σ_6, G_1, and δ_{00} we obtain a seminvariant whose leading term is free from γ_0.

We have shown in Articles VIII and XVI how to construct from a, Δ, S_3, σ_3, D, K, σ_5, σ_6, and δ_{00} seminvariants whose leading terms are $a^r \Delta^s \beta^u d^q$ $(r, s \geqq 0; u \geqq 1; q \geqq 1)$. Then subtracting from the remaining seminvariant the proper seminvariants constructed from a, Δ, S_3, σ_3, D, K, σ_5, σ_6, and δ_{00} we obtain a seminvariant whose leader is a polynomial in a and Δ. In Article XVI we have shown how to construct from a, Δ, G_1, σ_3, K, σ_7, and δ_{00} seminvariants with the same leading terms as all seminvariants whose leaders are $a^r \Delta^s$ $(r, s \geqq 1)$. Subtracting as before, we obtain a seminvariant whose leader is of the form $Aa^m + B\Delta^s$; the hypothesis of homogeneity shows that if $m = 1$, $B = 0$; but no such seminvariant can exist (vide IX supra). Therefore m is greater than 1 (if A is not zero). In Article XII we showed how to construct from a, S_3, D, and δ_{00} seminvariants led by $a^r d^q$ $(r \geqq 2)$. Subtracting as before we obtain a seminvariant whose leader is a numerical multiple of Δ^s. Directions for constructing from Δ, σ_3, K, and δ_{00} a seminvariant with the same leader as any existing seminvariant with such a leader have been given in Article XIII. Subtracting the proper seminvariant we have at last a seminvariant whose leading term involves no power of d higher than d^{q-1}.

Thus we have shown that by subtracting from any rational integral homogeneous modularly isobaric seminvariant S a seminvariant which is a polynomial in the thirteen seminvariants a, Δ, β, γ_0, σ_3, S_3, D, K, σ_5, σ_6, σ_7, G_1, and δ_{00} we may reduce S by at least one degree in d. By induction it follows that S is a polynomial in these thirteen seminvariants. Reducing β by the identity of XVI we have the following

THEOREM. *The twelve seminvariants* a, Δ, γ_0, S_3, D, σ_3, K, σ_5, σ_6, σ_7, G_1, *and* δ_{00} *are a fundamental system of seminvariants of the binary cubic form, modulo* 5.

C. A FUNDAMENTAL SYSTEM OF SEMINVARIANTS OF THE CUBIC, MODULO 7

XIX. SEMINVARIANTS LED BY a^r ($r \geqq 2$)

The case of seminvariants led by a, modulo p, has been treated in Article IX above. Seminvariants led by a^2 are the algebraic seminvariants S_3 and D, and the formal modular seminvariants

$$B_1 \equiv -\frac{1}{3}\sum_{t=0}^{6}(at^3+3bt^2+3ct+d)^5 \equiv a^2d^3+\cdots \pmod 7,$$

$$K \equiv -\sum_{t=0}^{6}(at^3+3bt^2+3ct+d)^6 \equiv a^2d^1+\cdots \pmod 7.$$

It follows from the theorem of Article VII that no seminvariants exist whose leading terms are a^2d^5 and a^2d^6; seminvariants whose leading terms are a^2d^{7r}, a^2d^{7r+1}, a^2d^{7r+2}, a^2d^{7r+3}, and a^2d^{7r+4} can be formed by the multiplication of the seminvariants S_3, D, B_1, and K by the proper power of δ_{00}; but the theorem of Article VII shows that no seminvariants exist whose leading terms are a^2d^{7r+5} and a^2d^{7r+6} ($r \geqq 1$). Seminvariants whose leading terms are a^3d^q ($q = 1, 2, 3, 4, 5, 6$) are aS_3, aD_6, aB_1, aK, and

$$B_2 \equiv -\sum_{t=0}^{6}(at^2+2bt+c)^3(at^3+3bt^2+3ct+d)^5 \equiv a^3d^5+\cdots \pmod 7,$$

$$B_3 \equiv -\sum_{t=0}^{6}(at^2+2bt+c)^3(at^3+3bt^2+3ct+d)^6 \equiv a^3d^6+\cdots \pmod 7.$$

Multiplying these and a^3 by the proper power of δ_{00} we have seminvariants led by a^3d^q ($q \geqq 7$) and multiplying these by a^{r-3} ($r \geqq 4$) we have seminvariants whose leading terms are a^rd^q ($r \geqq 4$; $q \geqq 7$).

We have thus shown that any existing seminvariant led by a or any higher power of a has the same leading term as a seminvariant which can be constructed from a, S_3, D, B_1, K, B_2, B_3, and δ_{00}.

XX. SEMINVARIANTS LED BY Δ^s ($s \geqq 1$)

There is a seminvariant led by Δ for every power q of d such that $q \equiv 0$, modulo 7, viz., $\Delta(\delta_{00})^r$ ($r \geqq 1$). For other powers of d there exist no seminvariants led by Δ. For a seminvariant led by Δ would be

$$\Delta d^q + Abc^2d^{q-1} + \cdots$$

and terms involving d^{q-1} after this is operated upon by Θ are

$$3q\Delta cd^{q-1} + Aac^2 d^{q-1} + 4Ab^2 cd^{q-1}.$$

In order that the supposed seminvariant be annihilated the sum of the coefficients of d^{q-1} must vanish, modulo 7. This requires that

$$3q + 4A \equiv 0, \text{ modulo } 7,$$

$$-3q + A \equiv 0, \text{ modulo } 7,$$

whence $A \equiv 0$, modulo 7. Applying the theorem of Article VII, we see that no such seminvariants exist.

Seminvariants whose leading terms are $\Delta^2 d^q$ $(q \geqq 1)$ are

$$C_1 \equiv -\tfrac{1}{3}\sum_{t=0}^{6}(at^2+2bt+c)^2(at^3+3bt^2+3ct+d)^3 \equiv \Delta^2 d + \cdots,$$

$$C_2 \equiv \sum_{t=0}^{6}(at^2+2bt+c)^2(at^3+3bt^2+3ct+d)^4 \equiv \Delta^2 d^2 + \cdots,$$

$$C_3 \equiv \tfrac{1}{4}\sum_{t=0}^{6}(at^2+2bt+c)^2(at^3+3bt^2+3ct+d)^5 \equiv \Delta^2 d^3 + \cdots,$$

$$C_4 \equiv -\sum_{t=0}^{6}(at^2+2bt+c)^2(at^3+3bt^2+3ct+d)^6 \equiv \Delta^2 d^4 + \cdots.$$

By multiplying these and Δ^2 by the proper powers of δ_{00}, we have seminvariants whose leading terms are $\Delta^2 d^q$ $(q \equiv 1, 2, 3, 4$, modulo 7$)$. That no seminvariants with the leading terms $\Delta^2 d^q$ $(q \equiv 5, 6$, modulo 7$)$ exist follows from the theorem of Article VII.

Seminvariants led by Δ^3 are $\Delta C_1, \Delta C_2, \Delta C_3, \Delta C_4$, and

$$C_5 \equiv a^3 B_2 - \sum_{t=0}^{6}(at^2+2bt+c)^6(at^3+3bt^2+3ct+d)^5 \equiv \Delta^3 d^5 + \cdots,$$

$$C_6 \equiv a^3 B_3 - \sum_{t=0}^{6}(at^2+2bt+c)^6(at^3+3bt^2+3ct+d)^6 \equiv \Delta^3 d^6 + \cdots.$$

By multiplying these and Δ^3 by $\Delta^{s-3}(\delta_{00})^t$ we obtain seminvariants whose leading terms are $\Delta^s d^{q+7t}$ $(s \geqq 4;\ t \geqq 1;\ q \geqq 1)$.

We have thus shown that any seminvariant led by Δ, Δ^2, or any higher power of Δ has the same leading term as a seminvariant which can be constructed from $\Delta, C_1, C_2, C_3, C_4, C_5, C_6$, and δ_{00}.

XXI. Seminvariants led by $a^r \Delta^s$ $(r, s \geqq 1)$

Seminvariants with the leader $a\Delta$ exist for every power q of d such that $q \equiv 0$, modulo 7, viz., $a\Delta(\delta_{00})^t$, $(t \geqq 1)$. For other powers of d no seminvariants led by $a\Delta$ exist. For such a seminvariant would be

$$a\Delta d^q + (Aabc^2 + Bb^3 c) d^{q-1} + \cdots$$

and the terms involving d^{q-1} after this is operated upon by Θ are

$$(3qab^2 c - 3qa^2 c^2 + Aa^2 c^2 + 4Aab^2 c + 3Bab^2 c + 2Bb^4) d^{q-1},$$

whence

$$\begin{aligned} 3q + 4A + 3B &\equiv 0, \text{ modulo } 7, \\ -3q + A \quad &\equiv 0, \text{ modulo } 7, \\ 2B &\equiv 0, \text{ modulo } 7, \end{aligned}$$

whence $A \equiv B \equiv 0$, modulo 7. This proves that no such seminvariants exist.

Seminvariants whose leading terms are $a^2 \Delta d^5$ and $a^2 \Delta d^6$, $a\Delta^2 d^5$ and $a\Delta^2 d^6$ are

$$E_1 \equiv \frac{1}{4}\sum_{t=0}^{6} (at^2 + 2bt + c)^4 (at^3 + 3bt^2 + 3ct + d)^5,$$

$$E_2 \equiv \frac{1}{4}\sum_{t=0}^{6} (at^2 + 2bt + c)^4 (at^3 + 3bt^2 + 3ct + d)^6,$$

$$E_3 \equiv \frac{1}{4}\sum_{t=0}^{6} (at^2 + 2bt + c)^5 (at^3 + 3bt^2 + 3ct + d)^5,$$

$$E_4 \equiv \frac{1}{4}\sum_{t=0}^{6} (at^2 + 2bt + c)^5 (at^3 + 3bt^2 + 3ct + d)^6.$$

By multiplying these together with S_3, D, K, B_1, C_1, C_2, C_3, and C_4, by proper powers of a, Δ, and δ_{00} we may show that any seminvariant led by $a^r \Delta^s$ has the same leading term as a seminvariant which can be constructed from a, Δ, S_3, D, B_1, K, C_1, C_2, C_3, C_4, δ_{00}, E_1, E_2, E_3, and E_4.

XXII. Seminvariants led by $a^r \Delta^s \gamma_0^t \beta^u$ $(r, s, t \geqq 0;\ u \geqq 1)$

We have shown in Article VIII above how to express βd^{q-1} $(q \geqq 1)$ as the sum of seminvariants whose leading terms are numerical multiples of $a^2 \Delta^{(p-3)/2} d^q$. For the modulus 7

$$\begin{aligned} \beta &\equiv 4\Delta^2 S_3 + 3a^2 C_1, \\ \beta d + \cdots &\equiv 2\Delta^2 D + 5a^2 C_2, \\ \beta d^2 + \cdots &\equiv 6\Delta^2 B_1 + a^2 C_3, \\ \beta d^3 + \cdots &\equiv \Delta^2 K + 6a^2 C_4, \\ \beta d^4 + \cdots &\equiv 3\Delta E_1 + 4aE_3, \\ \beta d^5 + \cdots &\equiv 5\Delta E_2 + 2aE_4, \\ \beta d^6 + \cdots &\equiv 6a^2 \Delta^2 \delta_{00} + KC_3. \end{aligned}$$

Proceeding as in the case of the modulus 5 we see that from a, Δ, γ_0, S_3, D, B_1, C_1, C_2, C_3, C_4, E_1, E_2, E_3, E_4, K, and δ_{00} we can construct seminvariants with the same leading terms as any seminvariant led by $a^r \Delta^s \gamma_0^t \beta^u$.

XXIII. Seminvariants led by $a^r \gamma_0^t$; seminvariants led by $\Delta^s \gamma_0^t$; seminvariants led by $a^r \Delta^s \gamma_0^t$

(1) This case may be treated exactly as was the case with seminvariants modulo 5, led by $a^r \gamma_0^t$ (vide XVII: 1 supra); the reader will notice that by Article VII no seminvariants exist whose leading terms are $a^2 \gamma_0^t d^{5+7w}$ and $a^2 \gamma_0^t d^{6+7w}$ $(w \geqq 0)$.

(2) No seminvariants exist whose leaders are $\Delta\gamma_0^t$ with the exception of $\Delta\gamma_0^t d^{7s} + \cdots$ $(s \geqq 0)$ (and these have the same leading terms as $\Delta\gamma_0^t(\delta_{00})^s$), for if such seminvariants existed they would be of the form

$$(\Delta c^{7t} + \text{terms of lower weight})\, d^q$$
$$+ (Abc^{7t+2} + \text{terms of lower weight})\, d^{q-1} + \cdots .$$

Operating on this supposed seminvariant with Θ we obtain

$$\{3q\Delta c^{7t+1} + Aac^{7t+2} + (14r + 4)Ab^2 c^{7t+1}\}d^{q-1}$$
$$+ (\text{terms of lower weight})\, d^q + \cdots$$

and as the sum of the terms of highest weight in the coefficient of d^{q-1} must be congruent to zero, modulo 7, we have

$$3q + (14r + 4)A \equiv 0, \text{ modulo } 7,$$
$$-3q + A \equiv 0, \text{ modulo } 7,$$

whence $A \equiv 0$, modulo 7, and the non-existence of the supposed seminvariant is proven, for by Article VII $Abc^{7t+2} d^{q-1}$ must not be zero if the supposed seminvariant exists.

C_v $(v = 1, 2, 3, 4)$ are seminvariants whose leading terms are numerical multiples of $\Delta^2 d^q$ $(q = 1, 2, 3, 4)$. Multiplication of these and $\Delta^2 \gamma_0^t$ by $(\delta_{00})^w$ gives us seminvariants whose leading terms are $\Delta^2 \gamma_0^t d^{7w+m}$ $(w \geqq 1;$ $m = 0, 1, 2, 3, 4)$. The theorem of Article VII again proves the non-existence of such seminvariants when $m = 5$ and 6. From the seminvariants whose leading terms are $\Delta^s d^q$ $(s \geqq 3;$ $q \geqq 1)$ already derived in Article XX we can by multiplication by $(\delta_{00})^n$ form seminvariants whose leading terms are $\Delta^s \gamma_0^t d^q$ $(s \geqq 3;$ $t \geqq 1;$ $q \geqq 1)$. We have thus shown how from Δ, γ_0, C_1, C_2, C_3, C_4, C_5, C_6, and δ_{00} to construct a seminvariant with the same leading term as any seminvariant led by $\Delta^s \gamma_0^t$.

(3) There exists for every value of q such that $q \equiv 0$, modulo 7, a semin-

variant led by a $\Delta\gamma_0'$, viz., $a\Delta\gamma_0'(\delta_{00})^r$ $(r \geqq 1)$. When $q \not\equiv 0$, modulo 7, no such seminvariant exists; for if there did it would be of the form

$$(a\Delta c^{7t} + \text{terms of lower weight})\,d^q$$
$$+ (Aabc^{7t+2} + Bb^3c^{7t+1} + \text{terms of lower weight})\,d^{q-1} + \cdots .$$

The terms of highest weight in the coefficient of d^{q-1} in the result of operating with Θ are

$$3qa\Delta c^{7t+1} + Aa^2c^{7t+2} + (14t+4)Aab^2c^{7t+1} + 3Bab^2c^{7t+1} + (14t+2)Bb^4c^{7t}.$$

Setting this congruent to zero, modulo 7, and proceeding as before we obtain inconsistent congruences in A and B.

We have shown in Article XXI how to construct from a, Δ, S_3, D, B_1, K, C_1, C_2, C_3, C_4, E_1, E_2, E_3, E_4, and δ_{00} seminvariants whose leading terms are $a^r\Delta^s d^q$ $(r, s \geqq 2;\ q \geqq 1)$. Multiplying these by γ_0' we have (together with those given above) seminvariants with the same leading terms as any seminvariants led by $a^r\Delta^s\gamma_0'$.

XXIV

The reader will now have no difficulty in seeing that by the method of Article XVIII it can be proved that *the twenty seminvariants* a, Δ, S_3, γ_0, D, K, B_1, B_2, B_3, C_1, C_2, C_3, C_4, C_5, C_6, E_1, E_2, E_3, E_4, *and* δ_{00} *are a fundamental system of seminvariants of the cubic form, modulo* 7.

That no one of these except γ_0 can be a polynomial in the others is evident from the fact that if we so multiply sets of them as to obtain a certain leading term in two ways the leader of the new seminvariant obtained by taking the difference of the two has in the most favorable case a leader of higher degree than the leader of any of the fundamental system except γ_0. Nor can the difference of any two such seminvariants be γ_0, for every term in the leader of the difference of two such seminvariants involves either a or b, whereas γ_0 has a term involving neither a nor b.

D. A FUNDAMENTAL SYSTEM OF FORMAL MODULAR PROTOMORPHS OF THE BINARY CUBIC, MODULO p

XXV

While one of the chief aims in the theory of algebraic seminvariants was the isolation of sets of seminvariants called fundamental, in terms of which every seminvariant could be expressed rationally and integrally, yet there have been discovered interesting sets of protomorphic seminvariants or protomorphs P_1, P_2, P_3, $\cdots$, P_n such that any seminvariant S of the form under consideration can be expressed in the form $A_1^q P(A_1, A_2, A_3, \cdots, A_n)$ where q

is a positive, negative or zero integer and P is a polynomial with integral coefficients in $A_1, A_2, A_3, \cdots, A_n$.

There is a corresponding theory of protomorphs in the formal modular seminvariant theory which has additional interest in the case of the binary cubic on account of the fact that while the number of members of a fundamental system of seminvariants of this form, modulo p, is a function of p, the number of protomorphs in a fundamental system is constant for any prime greater than 3. The set of protomorphs is also much simpler than the set of seminvariants, for it has only four members, and all of these save one are algebraic.

THEOREM. *The seminvariants a, Δ, S_3, and β form a set of protomorphs of the binary cubic, modulo p.*

Proof. Since

$$c = \frac{b^2 - \Delta}{a},$$

$$d = \frac{S_3 + 3abc - 2b^3}{a^2} = \frac{S_3 + b^3 - 3b\Delta}{a^2},$$

any seminvariant S of the cubic, modulo p, can be expressed as

$$\begin{aligned} S(a, b, c, d) &= S\left(a, b, \frac{b^2 - \Delta}{a}, \frac{S_3 + b^3 - 3b\Delta}{a^2}\right) \\ &= F\left(a, \frac{\Delta}{a}, \frac{S_3}{a^2}\right) + a^{-k} G(a, b, S_3, \Delta), \end{aligned}$$

where G is a polynomial in its arguments and F includes all the terms of S not involving b explicitly. Then G is divisible by b and hence by β. Treating the new seminvariant $H = G/\beta$ in like manner we see that $S = a^q P$, where q is a positive, negative, or zero integer, and P is a polynomial in a, Δ, S_3, and β.

Of the syzygies obtained, the following are examples, modulo 5:

$$\sigma_3 \equiv \frac{3\beta + \Delta S_3}{a^2}, \text{ modulo } 5;$$

$$\begin{aligned} \delta_{00} \equiv \frac{1}{a^{10}} (S_3^5 + 2a^4 \Delta S_3^3 + 3a^8 \Delta^2 S_3 + 4a^4 \Delta^4 S_3 + 4a^{12} S_3 \\ + 3a^4 S_3^2 \beta + 3\beta\Delta^5 + \beta^3 + 4a^4 \beta\Delta^3 + 2a^8 \beta\Delta), \text{ modulo } 5. \end{aligned}$$

WILLIAMSBURG, VIRGINIA,
February 5, 1920.

A PROPERTY OF TWO $(n+1)$-GONS INSCRIBED IN A NORM-CURVE IN n-SPACE*

BY

H. S. WHITE

§ 1. Introduction

Two cubic equations fix two triangles inscribed in a conic, if the coördinates of the generating point are given as quadric functions of one parameter. So for a gauche cubic curve, where homogeneous coördinates are given by rational integral functions of degree 3 in a parameter, two inscribed tetrahedrons may have the parameters of their vertices determined by any two quartic equations.

Two triangles inscribed in one conic determine a second conic which touches their six sides; and there exists a third conic with respect to which the two are reciprocal polar curves.† On a twisted cubic curve the analogous theorem still bears the name of von Staudt‡ as its originator, while Hurwitz has given its most accessible proof. It states that two tetrahedra inscribed in a gauche cubic determine uniquely a symmetric polarity in which they are self-reciprocal, and hence that their eight faces are osculating planes of a second gauche cubic curve.

Geometric proof of either theorem is not difficult, but a formula can be constructed which renders either one immediately visible. It is of interest to observe that the proof of the theorem simply as stated is most obvious if the formula is allowed to retain a certain extraneous factor; but the removal of this factor and the resulting condensation of the formula discloses more clearly the further fact that in each case the two sets of points employed are not unique, but are random selections from an infinite linear system of triads

* Presented to the Society, April, 1919, under different title.

† Brianchon, *Mémoire sur les lignes du second ordre.* Paris (1817), p. 35.

Steiner, *Die geometrischen Constructionen ausgeführt mittelst der geraden Linie und eines festen Kreises.* Berlin (1833), p. 67.

‡ See von Staudt, *Beiträge zur Geometrie der Lage*, p. 378; and A. Hurwitz, *Beweis eines Satzes aus der Theorie der Raumcurven III. Ordnung*, Mathematische Annalen, vol. 20 (1882), pp. 135–137. The latter establishes the existence of infinitely many such tetrahedra.

See also the *Encyklopädie der math. Wissenschaften*, III C 2, p. 236, § 108.

or quartettes. Moreover this method has an obvious extension to the norm-curve in flat space of n dimensions.

For each curve both primitive and reduced formulas will be exhibited.

§ 2. The conics and two triangles

On a conic the theorem cited amounts to asserting the existence of a $(2, 2)$ correspondence, symmetrical, among values of one parameter, which will convert each of three points into both the others, in each of two sets of three (or triads). Denote those parameters, in the two sets, by

$$a, b, c \qquad \text{and} \qquad a', b', c'.$$

If u is the original parameter, v the transformed, the following is a $(2, 2)$ correspondence or transformation which satisfies the requirements:

$$\begin{aligned}\phi(u, v) \equiv \\ &(c-b)\cdot(u-c)(u-b)\cdot(v-c)(v-b)\cdot(a-a')(a-b')(a-c') \\ &+(a-c)\cdot(u-a)(u-c)\cdot(v-a)(v-c)\cdot(b-a')(b-b')(b-c') \\ &+(b-a)\cdot(u-b)(u-a)\cdot(v-b)(v-a)\cdot(c-a')(c-b')(c-c')=0.\end{aligned}$$

For this relation is evidently satisfied identically by $u = a$, $v = b$, as each term contains either the factor $u - a$ or $v - b$, or both. Similarly for the pairs a, c and b, c. As for $u = a'$ and $v = b'$, insert those values in ϕ and remove the factor $(a' - a)(a' - b)(a' - c)\cdot(b' - a)(b' - b)(b' - c)$, whereupon the quotient remaining is the determinant

$$\begin{vmatrix} a - c' & b - c' & c - c' \\ 1 & 1 & 1 \\ a & b & c \end{vmatrix}$$

which vanishes.

This demonstrates the conic theorem, since when in $\phi(u, v)$ the quadric functions of u and v respectively are replaced by their equivalents in trilinear coördinates (x_1, x_2, x_3) and (y_1, y_2, y_3) of two points on the conic, $\phi(u, v)$ becomes symmetrically bilinear, and equated to zero gives a polar reciprocity

$$\phi(u, v) \equiv \Phi(x, y) = 0$$

with respect to a conic $\Phi(x, x) = 0$. Accordingly each vertex as a has its polar, as bc, touching the reciprocal of the conic on which the six vertices were taken to lie.

The extraneous factor in this formula $\phi(u, v)$ is

$$(a - b)(b - c)(c - a).$$

Remove that factor, and replace symmetric functions of either triad by the proper coefficient from one of these cubics:

$$f_1(u) = (u-a)(u-b)(u-c) = u^3 - Au^2 + Bu - C,$$
$$f_2(u) = (u-a')(u-b')(u-c') = u^3 - A'u^2 + B'u - C'.$$

We have then the reduced form

$$\begin{aligned}\phi_1(u,v) \equiv u^2v^2(A-A') - uv(u+v)(B-B') \\ + (u^2+uv+v^2)(C-C') + uv(AB'-A'B) \\ - (u+v)(AC'-A'C) - (BC'-B'C) = 0.\end{aligned}$$

Note that the constants are determinants from the array

$$\left\| \begin{matrix} 1 & A & B & C \\ 1 & A' & B' & C' \end{matrix} \right\|,$$

and that these are invariant save for the factor $(k_1l_2 - k_2l_1)$ when $f_1(u)$ and $f_2(u)$ are replaced by any two cubics

$$k_1f_1(u) + k_2f_2(u), \qquad l_1f_1(u) + l_2f_2(u)$$

of the linear system determined by the former two. Therefore all cubics of this linear system give polar triangles of the conic $\Phi(x,x) = 0$, and the sides of all such triangles touch one common curve of the second class.

§ 3. The twisted cubic and two tetrahedra

For the twisted cubic and the theorem of von Staudt and Hurwitz, the primitive formula is obviously the following,—summation covering cyclic permutations of a, b, c, d:

$$\sum \pm \begin{vmatrix} 1 & 1 & 1 \\ b & c & d \\ b^2 & c^2 & d^2 \end{vmatrix} \begin{aligned} &\cdot (u-b)(u-c)(u-d)\cdot(v-b)(v-c)(v-d) \\ &\cdot (a-a')(a-b')(a-c')(a-d') \equiv \phi(u,v) = 0.\end{aligned}$$

Point coördinates (x) and (y) upon the gauche cubic replace cubic expressions in u and v, giving a bilinear symmetric polarity

$$\phi(u,v) \equiv \Phi(x_1, x_2, x_3, x_4;\ y_1, y_2, y_3, y_4), \qquad \text{or} \qquad \Phi(x,y) = 0.$$

The quadric surface $\Phi(x,x) = 0$ is the one with respect to which the two tetrahedra are self-polar,—as the theorem asserts.

Here also, as in the preceding section, occurs a difference-product as an extraneous factor, and on its removal the polarity is seen to be covariant (or

combinant) in a linear system of tetrahedra. For if we denote by $f_1(u) = 0$ and $f_2(u) = 0$ the quartics whose roots are parameters of the vertices of the two tetrahedra,

$$f_1(u) \equiv (u-a)(u-b)(u-c)(u-d) \equiv u^4 - Au^3 + Bu^2 - Cu + D,$$

$$f_2(u) \equiv (u-a')\cdots(u-d') \equiv u^4 - A'u^3 + B'u^2 - C'u + D',$$

the function $\phi(u, v)$ can be represented as a determinant:

$$\phi(u,v)$$

$$\equiv \begin{vmatrix} f_2(a) & f_2(b) & \cdots & f_2(d) \\ (u-a)(v-a) & (u-b)(v-b) & \cdots & (u-d)(v-d) \\ a(u-a)(v-a) & b(u-b)(v-b) & \cdots & d(u-d)(v-d) \\ a^2(u-a)(v-a) & b^2(u-b)(v-b) & \cdots & d^2(u-d)(v-d) \end{vmatrix} = 0.$$

By a well-known theorem on alternants* this is reduced and the difference-product of a, b, c, d may be removed. The result, the essential form of the $(3, 3)$ relation, is the determinantal equation

$$\begin{vmatrix} D & C & B & A & 1 \\ D' & C' & B' & A' & 1 \\ uv & -(u+v) & 1 & 0 & 0 \\ 0 & uv & -(u+v) & 1 & 0 \\ 0 & 0 & uv & -(u+v) & 1 \end{vmatrix} = 0.$$

Thus either with or without the desirable symmetry of $\phi(u, v)$ in parameters of the first and second sets, *we have obtained a type-formula extensible at once to norm-curves in any number of dimensions.*

VASSAR COLLEGE,
April, 1919.

* Muir, *A treatise on the theory of determinants*, § 127.

RECURRENT GEODESICS ON A SURFACE OF NEGATIVE CURVATURE*

BY

HAROLD MARSTON MORSE

Introduction

The results necessary for the development of this paper are contained in a paper by G. D. Birkhoff,† in a paper by J. Hadamard,‡ and in an earlier paper by the present writer.§

In this earlier paper, as in the present paper, only those geodesics on the given surfaces of negative curvature are considered which, if continued indefinitely in either sense, lie wholly in a finite portion of space. A class of curves is introduced, each of which consists of an unending succession of the curve segments by which the given surface, when rendered simply connected, is bounded. It is shown how a curve of this class can be chosen so as to uniquely characterize some geodesic lying wholly in a finite portion of space. Conversely, it is shown that every geodesic lying wholly in a finite portion of space, is uniquely characterized by some curve of the above class.

The results of the earlier paper on geodesics, and the representation obtained there, will be used in the present paper to establish various theorems concerning sets of geodesics and their limit geodesics. In particular, the existence of a class of geodesics called *recurrent geodesics of the discontinuous type*,‖ will be established. This class of geodesics offers the first proof that has been given in the general theory of dynamical systems, of the existence of recurrent motions of the discontinuous type.

For a more complete treatment of the questions of the existence of surfaces of negative curvature, the reader is referred to the paper by Hadamard, already cited.

* Presented to the Society, Dec. 28, 1920.

† *Quelques théorèmes sur le mouvement des systèmes dynamiques*, Bulletin de la Société Mathématique de France, vol. 40 (1912), p. 303.

‡ *Les surfaces à courbures opposées et leur lignes géodésiques*, Journal de Mathématiques pures et appliquées, (5), vol. 4 (1898), p. 27.

§ *A one to one representation of geodesics on a surface of negative curvature*, American Journal of Mathematics, vol. 42 (1920).

‖ G. D. Birkhoff, loc. cit.

THE SURFACE

§ 1. We will consider surfaces without singularities in finite space. We will suppose the surface divisible into overlapping regions, such that every point of the surface lying in a finite portion of space is contained as an interior point in some one of a finite number of these regions, and such that the Cartesian coördinates x, y, z of the points of any one of these regions can be expressed in terms of two parameters, u and v, by means of functions with continuous derivatives up to a convenient order, at least the third, and such that

$$\left(\frac{D(xy)}{D(uv)}\right)^2 + \left(\frac{D(xz)}{D(uv)}\right)^2 + \left(\frac{D(yz)}{D(uv)}\right)^2 \neq 0.$$

By a *curve* on the surface we will understand any set of points on the surface in continuous correspondence with the points of an interval on a straight line, including one, both, or neither of its end points.

We will suppose the Gaussian curvature of the surface to be negative at every point, with the possible exception of a finite number of points, at which points the curvature will necessarily be zero. A first result, given by Hadamard in the paper already referred to, is that a surface of negative curvature cannot be contained in any finite portion of space.

§ 2. By a *funnel* of a surface will be meant a portion of a surface topographically equivalent to either one of the two surfaces obtained by cutting an unbounded circular cylinder by a plane perpendicular to its axis. We will consider surfaces of negative curvature whose points, outside of a sufficiently large sphere with center at the origin, consist of a finite number of funnels. Each of these funnels will be cut off from the rest of the surface along a simple closed curve. These curves will be taken sufficiently remote on the funnels to be entirely distinct from one another.

An unparted hyperboloid of revolution is an example of a surface of negative curvature with two funnels.

From the definition of a funnel it follows that, by a continuous deformation of the closed curve forming the boundary of the funnel, the funnel may be swept out in such a way that every point of the funnel is reached once and only once. Hadamard considers two classes of funnels: those which can be swept out by closed curves which remain less in length than some fixed quantity, and those which do not possess this property. Surfaces with funnels of the first sort are for several reasons of less general interest than those with funnels of the second sort. In the present paper surfaces with funnels only of the second sort will be considered. Hadamard showed that there exist surfaces of negative curvature possessing funnels all of the second sort, of any arbitrary number exceeding one, and such that the surface obtained by cutting off these funnels is of an arbitrary genus.

§ 3. We shall consider surfaces which possess at least two funnels of the second sort, and of the surfaces with just two funnels of the second sort, we will exclude those surfaces that are topographically equivalent to an unbounded circular cylinder. Hadamard proves that on such surfaces there exists one and only one closed geodesic that is deformable into the boundary of a given funnel, and that this geodesic possesses no multiple points, and no points in common with the other closed geodesics that are deformable into the boundaries of the other funnels.

We shall denote these closed geodesics, say v in number, by

$$g_1\, g_2 \cdots g_v . \tag{1}$$

They will form the complete boundary of a part of the surface, contained in a finite part of space. We denote this bounded surface by S. As shown in § 18 and § 19 of my earlier paper on geodesics, S may be rendered simply connected as follows: S is first cut along a system of geodesics,

$$h_1\, h_2 \cdots h_{v-1},$$

each of which has one end point on an arbitrarily chosen point, P, on g_v, and the other, respectively, on the geodesic of the set

$$g_1\, g_2 \cdots g_{v-1},$$

with the same subscript, and no two of which have a point other than P in common, and no points other than their end points in common with the geodesics, of the set (1). There then results a surface with a single boundary. This surface can be rendered simply connected by $2p$ geodesics,

$$c_1\, c_2 \cdots c_{2p},$$

which can be taken as beginning and ending at P, and which will have no other points than P in common with any of the other geodesics or with each other.

We denote by T, the simply connected piece of surface obtained by cutting S along the above geodesics. It may be proved as a consequence of the assumptions made concerning the representation of the given surface, that T is topographically equivalent to a plane region consisting of the interior and boundary points of a circle.

Representation of geodesics by linear sets or by reduced curves

§ 4. We suppose that we have at our disposal an unlimited number of copies of the simply connected surface T, and that each of these copies of T is entirely distinct from every other copy of T.

Definition. Let r be any integer, positive, negative or zero. Let T_r

denote a particular copy of T. By a *linear set* of copies of T will be understood a surface consisting of a set of copies of T of the form

$$(1) \qquad \cdots T_{-2}\, T_{-3}\, T_0\, T_1\, T_2 \cdots ,$$

or of the form of any subset of successive symbols of (1), in which no one copy of T appears twice, and in which each copy of T is joined along some one of its boundary pieces to that boundary piece of the succeeding copy of T that arises from the opposite side of the same cut, while no copy of T is joined to its predecessor and successor along the same boundary piece. A linear set which has no first or last copy of T will be called an *unending* linear set.

Two linear sets will be considered the *same* if the two sets of their copies of T can both be expressed by the same form (1), in such a manner that successive symbols represent successive copies of T in the respective linear sets, joined along copies of the *same* cut.

A linear set in which the number of copies of T is finite is seen to be a multiple-leaved, simply connected surface, bounded by a single closed curve.

Let the set of geodesic segments,

$$g_1\, g_2 \cdots g_v;\ h_1\, h_2 \cdots h_{v-1};\ c_1\, c_2 \cdots c_{2p},$$

described in § 3, be denoted by H.

Definition. Let r be any integer, positive, negative, or zero. Let k_r be any member of the set H. By a *reduced curve* we shall understand any *continuous* curve that consists of a set of members of the set H, excluding g_v, of the form

$$(1) \qquad \cdots k_{-2}\, k_{-1}\, k_0\, k_1\, k_2 \cdots ,$$

or of the form of any subset of consecutive symbols of (1). In the special* case where a k_r and a k_{r+1} of (1) are copies of the same member of the set H, say l, we require that the end point of k_r and the end point of k_{r+1} which are joined, be points which on l would be considered as opposite end points. A reduced curve without end points will be termed an *unending* reduced curve.

If a given reduced curve be traced out in an arbitrary sense, it follows from the last condition of the definition of a reduced curve that no two consecutive pieces of the given reduced curve will thereby appear as copies of the same piece of H taken in opposite senses.

§ 5. The results of this section are established in § 12 and § 13, of my earlier paper on geodesics, already referred to.

A given unending reduced curve is contained in one and only one linear

* We admit the possibility of two symbols in (1) representing the same member of the set H, but as parts of the reduced curve we shall consider two such copies as distinct, in a manner analogous to the convention ordinarily made in the construction of a Riemann surface.

set, which set is an unending linear set. Conversely, every unending linear set contains one and only one unending reduced curve. Each copy of T of an unending linear set that contains an unending reduced curve contains either a point or a single continuous segment of the given reduced curve, and no other points of the given reduced curve. The results necessary for the developments of this paper are summed up in the following:

THEOREM 1. *There is a one to one correspondence between the set of all unending reduced curves on S, and the set of all unending linear sets, in which each reduced curve corresponds to that linear set in which it is contained.*

§ 6. The results of this section follow from the results of §§ 21, 22, and 23, of the earlier paper on geodesics. For the purpose of representing geodesics that lie wholly on S, it will be convenient to suppose each closed geodesic replaced by that geodesic obtained by tracing out the given closed geodesic an infinite number of times in either sense.

Every geodesic lying wholly on S is contained on one and only one linear set, which set must be an unending linear set. Conversely, every unending linear set contains one and only one of the geodesics lying wholly on S. Every copy of T of a linear set that contains a geodesic lying wholly on S, contains either a point or a single continuous segment of this geodesic, and no other point on this geodesic.

THEOREM 2. *There is a one to one correspondence between the set of all geodesics lying wholly on S, and the set of all unending linear sets, in which each geodesic corresponds to that linear set in which it is contained.*

The results of Theorems 1 and 2 can be combined in the following:

THEOREM 3. *There is a one to one correspondence between the set of all geodesics lying wholly on S, and the set of all unending reduced curves, in which each geodesic corresponds to that unending reduced curve that is contained in the same linear set.*

THEOREM 4. *If an unending reduced curve consists wholly of repetitions of a closed curve, the geodesic that passes through the same linear set consists wholly of successive repetitions of a closed geodesic. Conversely, if a geodesic consists wholly of successive repetitions of a closed curve, the unending reduced curve that passes through the same linear set, consists wholly of successive repetitions of a closed curve.*

VARIATION OF GEODESICS WITH INITIAL ELEMENTS

§ 7. On a surface representable in the manner in which the given surface is representable, there is one and only one geodesic through a given point, and tangent to a given direction.

DEFINITION. A point on the surface and a direction tangent to the surface will be called an *element*, and will be said to define that sensed geodesic that

passes through the initial point of the given element, and is such that its positive tangent direction at that point agrees with the direction of the given element.

If u and v are parameters in any representation of a part of the surface, and if θ' is the angle which a given tangent direction makes at the point $(u' v')$ with the positive tangent to the curve, $u = u'$, then $(u' v' \theta')$ will represent an element of the given surface. We shall understand by each statement of metric relations between elements, the same statement of metric relations between the points in space of three dimensions obtained by considering the complex $(u' v' \theta')$ as the Cartesian coördinates of a point.

Let G be any geodesic segment lying on the original uncut surface. G is an extremal in the Calculus of Variations problem of minimizing the arc length, from which theory we can readily obtain the following theorem that describes the nature of the variation of G with variation of its initial element.*

THEOREM 5. *Corresponding to any positive constants e and h, there exists a positive constant d so small, that if any two elements, with initial points on the bounded surface S, lie within d of each other, and if a second pair of elements lie respectively on the two geodesics defined by the first two elements, and if further the initial points of this second pair of elements lie respectively at a distance, measured along the given geodesics from the geodesics' initial points, that is the same in both cases and that does not exceed h, the second pair of elements will lie within e of each other.*

The following theorem describes the manner in which a geodesic varies with the reduced curve contained in the same linear set. It is given in § 24 of the earlier paper on geodesics.

THEOREM 6. *Corresponding to any positive constant e, there exists a positive constant k, so large, that if two unending reduced curves possess in common a continuous segment of length exceeding k, the two corresponding geodesics each have at least one element within e of some element on the other, and with initial point in the same copy of T, in the geodesic's linear set, as the mid point of the common reduced curve segment.*

Conversely, corresponding to any positive constant k, there exists a positive constant e, so small, that if on each of two geodesics there exists some element within e of some element on the other, the two corresponding reduced curves possess in common a segment of length k, with mid point in the same copy of T in the reduced curve's linear set, as the initial point of either of the two elements.

REPRESENTATION OF GEODESICS BY SETS OF NORMAL CURVES

§ 8. The previous representation of geodesics by means of linear sets and reduced curves can now be replaced by another representation which will be

* Cf. Bolza, *Vorlesungen über Variationsrechnung* (1909), p. 219.

fundamental in the work of this paper. This representation will be in terms of the geodesic segments,

(1) $$c_1 c_2 \cdots c_{2p},$$

(2) $$g_1 g_2 \cdots g_{v-1},$$

which form a subset of the boundary pieces of each copy of T.

DEFINITION. I. Each one of the geodesic segments of (1) and (2) will be called a *normal segment.*

II. Let m be any integer, positive, negative, or zero. Let C_m represent any sensed normal segment. By a *normal set* C, will be understood an unending ordered set of sensed normal segments, in the form

(3) $$\cdots C_{-2} C_{-1} C_0 C_1 C_2 \cdots,$$

in which no two successive members are the same normal segment taken in opposite senses.

III. Two normal sets C will be considered the *same* if they contain the same normal segments in the same order with the same senses.

A normal set C will not in general constitute a reduced curve. For a reduced curve may include any normal segment, and in addition any geodesic segment of the set,

(4) $$h_1 h_2 \cdots h_{v-1}.$$

However, it is readily seen that, with the aid of the members of the set (4), there can be formed from a given normal set C one and only one sensed reduced curve whose normal segments taken in the order and with the senses in which they appear on the given reduced curve constitute the given normal set C. Conversely, if there be given any unending sensed reduced curve, its normal segments taken in the order and with the senses in which they appear on the given unending sensed reduced curve, constitute a normal set C.

Thus there is a one to one correspondence between the set of all unending sensed reduced curves and the set of all normal sets C, in which each normal set C corresponds to that unending sensed reduced curve whose normal segments, taken in the order and with the senses in which they appear on the given unending sensed reduced curve, constitute the given normal set C.

DEFINITION. If an unending sensed reduced curve and a normal set C correspond in the sense of the preceding statement, the normal set C will be said to *represent* the given unending sensed reduced curve, and also that sensed geodesic that passes through the same linear set of copies of T in the same sense as does the given unending sensed reduced curve.

By virtue of Theorem 3, § 6, every sensed geodesic lying wholly on the surface S, is represented by one and only one normal set C, while every normal set C represents one and only one sensed geodesic.

Closed geodesics

§ 9. If a normal set C of the from (3) of the preceding section represents a closed geodesic it follows from Theorem 4, § 6, that there exists a positive integer p, such that in the set C

$$C_m = C_{m+p},$$

where m is any integer, positive, negative, or zero. The given normal set will then be said to be *periodic*, and to have the period p. Theorem 4, § 6 now becomes the following:

Theorem 7. *A necessary and sufficient condition that a geodesic be closed, is that the normal set C representing that geodesic be periodic.*

Let q be the smallest period of a periodic set C. Then any other period p must either equal q, or else be a multiple of q. For if p were not equal to q or a multiple of q, it follows from Euclid's Algorithm that there exist three integers, A, B, and r, of which r is less than q, and is greater than zero, and which are such that

$$Aq + Bp = r.$$

It follows from this equation that r is also a period of the given periodic set C, contrary to the assumption that q was the smallest period of the given periodic set C.

Definition. If q is the smallest period of a periodic normal set C, then any q successive sensed normal segments of the given set C will be called a *generating set* of the given set C, and also of the closed geodesic represented by the given set C.

If B is a generating set of a normal set C, this set C consists merely of an unending succession of sets B, which we will write in the form,

$$\cdots B\,B\,B\,B\,B\,B\,B\,B \cdots$$

All generating sets of a periodic set C, can evidently be obtained from any one such generating set by a circular permutation of the sensed normal segments composing the given generating set.

§ 10. We consider now the question of the arbitrary formation of sets that may serve as generating sets of some geodesic. To that end we form a finite ordered set of sensed normal segments, in which neither the first and last members, nor any two successive members are the same normal segment taken in opposite senses, and which cannot be obtained through repetitions of a similar set containing fewer sensed normal segments. Denote the set so obtained by D. The set

$$\cdots D\,D\,D\,D\,D \cdots \tag{1}$$

is, in the first place, a normal set C. For D is made up of sensed normal

segments in which neither the first and last members, nor any two successive members are the same normal segment taken in opposite senses. Further D is a generating set of the set (1), for otherwise (1) would have a period smaller than the number of successive segments in D, and hence a period that is a divisor of the number of successive segments in D. D could then be obtained by a finite number of repetitions of a similar set containing fewer sensed normal segments, contrary to the last hypothesis made concerning D.

The number of different periodic normal sets C is seen to equal the number of generating sets not obtainable one from the other by a circular permutation of their normal segments. The number of such generating sets is readily seen to be an enumerable infinity. From this result, together with the theorem of the preceding section, we have the result given by Hadamard:

There are an enumerable infinity of distinct closed geodesics on the surface S.

Limit geodesics of sets of geodesics

§ 11. Definition. A geodesic G will be said to be a *limit geodesic of a set of geodesics* if a set of elements, M, lying on the given set of geodesics, have as a limit an element E, on G, while all the initial points of those elements of the set M that lie on G, are at distances, measured along G from the initial point of E, exceeding a fixed positive quantity.

From the property of continuous variation of a geodesic with its initial element, as given in Theorem 5, § 7, it follows that if one element on G is a limit element of elements on a given set of geodesics, then every element on G is a limit element of elements on the given set of geodesics.

If a closed geodesic should be considered as replaced by an unclosed geodesic that traces out the given closed geodesic an infinite number of times in either sense, the latter geodesic would be a limit geodesic of itself. In this sense any closed geodesic will be considered a limit geodesic of itself.

From Theorem 6, § 7, it follows that a necessary and sufficient condition that a geodesic G be a limit geodesic of a set of geodesics J, not including G, is that every finite segment of the unending reduced curve corresponding to G be contained in the unending reduced curve corresponding to some geodesic of the set J. In terms of normal sets C, this result becomes the following:

Theorem 8. *A necessary and sufficient condition that a geodesic G be a limit geodesic of a set of geodesics J, not including G, is that every subset of consecutive normal segments of the normal set representing G, be a subset of consecutive normal segments of some normal set representing a geodesic of the set J.*

The following theorem is given by Hadamard, with a proof, however, that is different from the following.

Theorem 9. *Every geodesic lying wholly on a surface of negative curvature*

for which $2p + v - 1 \geqq 2$ (cf. section 3), *is a limit geodesic of the set of all closed geodesics on that surface.*

The number of different normal segments equals $2p + v - 1$. Hence on any of the surfaces considered, there are at least two different normal segments. Since any closed geodesic is a limit geodesic of itself, we need only consider the case of a geodesic not a closed geodesic. Let G be any geodesic lying wholly on S, and not a closed geodesic. Let there be given an arbitrary finite subset of consecutive normal segments of the normal set C representing G. If this subset does not begin and end with the same normal segment taken in opposite senses, we denote the subset by D; in the other case we add to the given subset a normal segment different from the first and last normal segment, and denote this set also by D.

In either case,

$$\cdots D\,D\,D\,D\,D\,D\,D \cdots$$

will be a normal set C. This normal set is periodic; according to Theorem 7, § 9, it then represents a closed geodesic. Further this normal set contains as a subset of successive normal segments the given arbitrary subset of the normal set representing G. From Theorem 8, it accordingly follows that G is a limit geodesic of the set of all closed geodesic on S, and the theorem is proved.

THEOREM 10. *On a surface of negative curvature for which* $2p + v - 1 \geqq 2$, *there exists at least one geodesic which has for a limit geodesic every geodesic lying wholly on* S.

The set of all possible finite subsets of consecutive normal segments of normal sets C, form an enumerable set which may accordingly be put into one to one correspondence with the set of all integers, positive, negative, or zero. In this correspondence that one of these subsets that corresponds to the integer n, we denote by B_n. The set,

$$\cdots B_{-2}\,B_{-1}\,B_0\,B_1\,B_2 \cdots$$

will be a normal set C, unless for some integer n, the last sensed normal segment of B_{n-1} and the first sensed normal segment of B_n are the same normal segments taken in opposite senses. In every such case we insert between B_{n-1} and B_n a normal segment different from the normal segment in question. The resulting set will be a normal set, which we denote by C'.

C' contains each subset B_n as a subset of consecutive normal segments. It follows from Theorem 8, that every geodesic lying wholly on S, with the possible exception of the geodesic represented by C', is a limit geodesic of the geodesic represented by C'. That the geodesic represented by C' is a limit geodesic of itself, follows from the fact that every *closed* geodesic is a

limit geodesic of the geodesic represented by C', while *every* geodesic lying wholly on S is a limit geodesic of the set of all closed geodesics on S.

Recurrent geodesics

§ 12. The following definition, and Theorems 11, 12, and 13, are restatements for the case of geodesics of what is given by Professor Birkhoff for a dynamical system, in the paper referred to in the introduction.

Definition. By a *minimal set* of geodesics we shall understand any set of geodesics lying wholly on S, each of which has every other geodesic of the set, and no other geodesic, as a limit geodesic. Any geodesic of a minimal set will be called a *recurrent* geodesic.

A closed geodesic constitutes a minimal set in which it is the only geodesic.

The following theorem serves as an existence proof for recurrent geodesics.

Theorem 11. *Every geodesic lying wholly on S contains among its limit geodesics at least one minimal set of geodesics.*

Concerning the number of recurrent geodesics in a minimal set, we have

Theorem 12. *The power of any minimal set not simply a closed geodesic, is that of the continuum.*

The characteristic property of a recurrent geodesic is given by the following:

Theorem 13. *A necessary and sufficient condition that a geodesic lying wholly on S be a recurrent geodesic is that, corresponding to any arbitrary positive constant e, there exist a positive constant h, so large, that if L be any segment of the given geodesic of length at least equal to h, any element of the given geodesic lies within e of some element of L.*

§ 13. Let there be given a set of symbols of the form,

$$\cdots R_{-2}\, R_{-1}\, R_0\, R_1\, R_2 \cdots . \tag{1}$$

Let m and n be any integers, positive, negative, or zero.

Definition. I. A set of symbols of the form (1) will be said to be *recurrent*, if corresponding to any positive integer r, there exists a positive integer s, so large that any subset of (1) of the form,

$$R_m\, R_{m+1} \cdots R_{m+r} \tag{2}$$

is contained in every subset of (1) of the form

$$R_n\, R_{n+1} \cdots R_{n+s} . \tag{3}$$

II. The set (1) will be said to be *periodic*, if there exists a positive integer p, such that

$$R_n = R_{n+p},$$

whatever integer n may be, and p will be said to be a *period* of the set (1).

It appears at once that a set (1) that is periodic, is also recurrent.

Theorem 13, § 12, interpreted in terms of normal sets C by means of Theorem 6, § 7, becomes the following

THEOREM 14. *A necessary and sufficient condition that a geodesic lying wholly on S be recurrent, is that the set C representing the given geodesic be recurrent.*

EXISTENCE OF RECURRENT GEODESICS, NOT PERIODIC.

§ 14. We come now to the question of the existence of recurrent geodesics that are not closed geodesics.

On a surface of negative curvature topographically equivalent to an unbounded circular cylinder, the only possible recurrent geodesic is a single closed geodesic. On a surface of negative curvature topographically equivalent to an unbounded plane, there are no recurrent geodesics whatever. The surfaces of negative curvature which we have been considering include neither of these two types of surfaces (cf. § 3).

We have seen in Theorem 7, § 9, that a geodesic that is periodic is represented by a normal set C that is periodic; while Theorem 14, § 13, states that a normal set C that is recurrent represents a geodesic that is recurrent. Hence, to prove the existence of a geodesic that is recurrent without being periodic, it is sufficient to prove the existence of a normal set C that is recurrent without being periodic.

Now there are just $2p + v - 1$ normal segments (cf. § 8). We are considering surfaces for which $2p + v - 1 \geqq 2$. Hence any of the surfaces of negative curvature considered will possess at least two normal segments. We will seek a normal set C that is composed solely of two normal segments. For that purpose the following lemma is introduced.

LEMMA. *There exists an unending set of symbols each of which is either* 1 *or* 2, *which forms a set that is recurrent without being periodic.*

By the juxtaposition of two or more symbols representing ordered sets of symbols, we shall mean here, as elsewhere, the ordered set obtained by taking the symbols of the given sets in the order in which the sets are written.

Let n be any positive integer. We introduce the following definitions:

$$\begin{aligned} a_0 &= 1, \\ b_0 &= 2, \\ a_1 &= a_0 b_0, \\ b_1 &= b_0 a_0, \\ &\cdots \\ a_{n+1} &= a_n b_n, \\ b_{n+1} &= b_n a_n. \end{aligned} \tag{1}$$

We introduce the set of symbols,

(2) $$\cdots d_{-2}\, d_{-1}\, d_0\, d_1\, d_2 \cdots,$$
of which

$$d_0\, d_1 \cdots d_{2^n}$$

are defined respectively as the 2^n integers of a_n; further, if m is any positive integer, d_{-m} is defined as equal to d_{m-1}. The set (2), so defined, will be proved to be recurrent without being periodic.

For definiteness we write out (2) in part, beginning with d_0:

(3) 1221 2112 2112 1221 2112 1221 $\cdots$.

It follows from the definitions (1), that if the integers of (2) be grouped in groups of 2^n integers, then the set (2) can be expressed, beginning with d_0, by a succession of the sets a_n and b_n, obtained by replacing the integers 1 and 2 in the set (2), respectively by a_n and b_n. Thus beginning with d_0, (2) is given in part as

(4) $$a_n\, b_n\, b_n\, a_n \quad b_n\, a_n\, a_n\, b_n \quad b_n\, a_n\, a_n\, b_n \quad \cdots.$$

The symbols of the set (2) that have negative subscripts, can be obtained, according to their definition, by taking the symbols of (2) with positive or zero subscripts in reverse order. It follows from the definitions (1) that the integers of a_n and b_n, taken respectively in their reverse orders, give a_n and b_n when n is even, and b_n and a_n when n is odd. We have the result:

Whatever integer n may be, the set (2) can be expressed by a properly chosen succession of the sets, a_n and b_n. Thus, if r be any integer such that

$$r \equiv 0 \textit{ modulo } 2^n,$$

then any subset of (2) of the form

$$d_r\, d_{r+1} \cdots d_{r+2^n}$$
is either a set a_n or a set b_n.

We will now prove that the set (2) is recurrent.

Let there be given any subset of (2) of the form

(5) $$d_s\, d_{s+1} \cdots d_{s+m},$$

where s is any integer, positive, negative, or zero, and m is any positive integer. Let r' be the largest integer less than s such that

$$r' \equiv 0 \text{ modulo } 2^m.$$

From the choice of r', we have,

$$r' < s < s + m < r' + 2^{m+1}.$$

Hence the set (5) is a subset of the set,

$$d_{r'}\, d_{r'+1} \cdots d_{r'+2^{m+1}}. \tag{6}$$

From the result of the preceding paragraph, it appears that (6) must be one of the four possible ordered combinations of a_m and b_m, that is, one of the four sets,

$$a_m\, b_m\,, \quad b_m\, b_m\,, \quad b_m\, a_m\,, \quad a_m\, a_m\,.$$

Each of these four sets is a subset of a_{m+3} and b_{m+3}; for from the equations (1), we have,

$$a_{m+3} = a_{m+2}\, b_{m+2} = a_{m+1}\, b_{m+1}\, b_{m+1}\, a_{m+1} = a_m\, b_m\, b_m\, a_m\, b_m\, a_m\, a_m\, b_m\,,$$

$$b_{m+3} = b_{m+2}\, a_{m+2} = b_{m+1}\, a_{m+1}\, a_{m+1}\, b_{m+1} = b_m\, a_m\, a_m\, b_m\, a_m\, b_m\, b_m\, a_m\,.$$

Since the set (2) can be expressed as a succession of the sets a_{m+3} and b_{m+3}, each of which contains 2^{m+3} integers of the set (2), it appears that any subset of at least 2^{m+4} successive integers of (2), say R, contains at least one of the sets a_{m+3} and b_{m+3}. Retracing the steps it is seen that R contains a subset identical with the given set (5). The set (2) is thus recurrent.

We will now show that the set (2) is not periodic. Suppose that the set (2) had a period prime to 2. Since p is prime to 2, there exists an integer m, greater than one, such that

$$2 \equiv 2^m, \qquad \text{modulo } p. \tag{7}$$

Since the set (2) has the period p, it follows from (7) that the set

$$d_2\, d_3\, d_4 \cdots \tag{8}$$

must be identical with the set

$$d_{2^m}\, d_{2^m+1}\, d_{2^m+2} \cdots. \tag{9}$$

The set (8) commences with the integers

$$2\ 1\ 2\ 1\ 1\ 2 \cdots, \tag{10}$$

while the set (9) commences as does b_m, which is seen from the equations (1) to commence with the integers

$$2\ 1\ 1\ 2\ 1\ 2 \cdots. \tag{11}$$

The sets (10) and (11) are not identical. The set (2) can thus have no period prime to p.

Finally suppose that the set (2) had a period $2^r\, p$, where r is any positive integer, and p is prime to 2. Let the set (2), commencing with d_0, be written in terms of a_r and b_r:

$$a_r\, b_r\, b_r\, a_r\, b_r\, a_r \cdots. \tag{12}$$

Considered as a succession of symbols a_r and b_r, (12) has the period p. But the original expression for (2) in terms of its integers, and commencing with d_0, is obtained from (12) by replacing the symbols a_r and b_r respectively by 1 and 2. Thus the expression for the set (2), in terms of its integers, and commencing with d_0, would have a period prime to 2. We have seen this to be impossible.

Thus the set (2) is recurrent without being periodic, and the lemma is proved.

§ 15. THEOREM 15. *On a surface of negative curvature for which*

$$2p + v - 1 \geqq 2,$$

there exists a set of geodesics that are recurrent without being periodic, and this set has the power of the continuum.

The number of different normal segments equals $2p + v - 1$. We are considering surfaces of negative curvature for which $2p + v - 1 \geqq 2$. Hence on any of the surfaces considered, there are at least two different normal segments. Let N_1 and N_2 be two different normal segments, each taken in an arbitrary sense.

In the preceding lemma we have established the existence of a set,

$$\cdots d_{-2}\, d_{-1}\, d_0\, d_1\, d_2 \cdots,$$

that is composed entirely of the integers one and two, and which is a set that is recurrent without being periodic. The set

$$\cdots N_{d_{-2}}\, N_{d_{-1}}\, N_{d_0}\, N_{d_1}\, N_{d_2} \cdots \tag{1}$$

is accordingly recurrent; from Theorem 14, § 13, it follows that the geodesic represented by the set (1) is recurrent. The set (1) is not periodic; it follows from Theorem 7, § 9, that the geodesic represented by (1) is not periodic. We have thus established the existence of a geodesic that is recurrent without being periodic.

According to Theorem 12, § 12, the existence of one geodesic that is recurrent without being periodic, is sufficient to establish that the power of the complete set of geodesics that are recurrent without being periodic is that of the continuum.

§ 16. THEOREM 16. *On a surface of negative curvature for which*

$$2p + v - 1 \geqq 2,$$

the set of all geodesics that are recurrent without being periodic, has as a limit geodesic every geodesic lying wholly on S.

Let there be given an arbitrary closed geodesic lying on S. Let B be any finite subset of successive normal segments of the normal set C representing the given closed geodesic. Let N_1 and N_2 be the two sensed normal segments

used in the proof of the preceding theorem, and $-N_1$ and $-N_2$ be, respectively, the same normal segments taken in opposite senses.

If now the set B does not begin or end with $-N_1$, or $-N_2$, we denote the set B, by D. If the set B begins with $-N_1$ or $-N_2$, we prefix N_2 or N_1, respectively, to the set B, while if the set B ends with $-N_1$ or $-N_2$, we add N_2 or N_1, respectively to the set B, and in either case denote the resulting set by D. We interpose this set D between each two successive sensed normal segments of the normal set C, given by (1) in the proof of the preceding theorem, and denote the resulting set by C'.

It is a consequence of the nature of the construction of the set C', that no two of its successive sensed normal segments are the same normal segment taken in opposite senses. The set C' is thus a normal set. The normal set (1) of the proof of the preceding theorem, is recurrent without being periodic; it follows that the set C' is recurrent without being periodic. The geodesic represented by C' is accordingly recurrent without being periodic. The set C' contains B as a subset of successive normal segments. It follows from Theorem 8, § 11, that the given closed geodesic is a limit geodesic of the set of all recurrent geodesics that are not periodic.

That every geodesic lying wholly on S is a limit geodesic of the set of all geodesic that are recurrent without being periodic, follows now from the fact that *every* geodesic lying wholly on S is a limit geodesic of the set of all closed geodesics on S.

Distribution of elements on recurrent geodesics

§ 17. Definition. Two elements E' and E'' of a set of elements M on a region R of S, will be said to be *mutually accessible in M and on R*, if corresponding to any positive constant e, there exists in the set M, a finite ordered subset of elements of which the first is E', and the last E'', while each element of the subset, excepting the last, lies within a geodesic distance e, measured on R, of the following element.

The following theorem is established in § 25 of the earlier paper on geodesics.

Theorem 17. *On any simply connected region R of S, and in the set of all elements on R, and on geodesics lying wholly on S, no two elements on different geodesics are mutually accessible.*

A particular consequence of the preceding theorem is that, on any simply connected region R of S, and in the set of all elements on R, and on geodesics that are *recurrent*, no two elements on different geodesics are mutually accessible. A set of recurrent geodesics with this property are of a type called *discontinuous* recurrent motions by Professor Birkhoff. Thus:

Theorem 18. *The set of all recurrent geodesics on S constitutes a set of recurrent motions of the discontinuous type.*

The proof, given in this paper, of the existence of a set of this type, is the first proof of the existence of a discontinuous set of recurrent motions.

§ 18. A recurrent geodesic was defined as a member of a minimal set,—a set in which every geodesic has every geodesic of the set, and no other geodesic, as a limit geodesic. In case a given recurrent geodesic is a closed geodesic, the minimal set containing the given geodesic consists merely of the given closed geodesic. In case a recurrent geodesic is not a closed geodesic, the power of the minimal set that contains the given recurrent geodesic, is, according to Theorem 12, § 12, that of the continuum.

From the definition of a minimal set, it appears that no two minimal sets that are not identical, have any geodesic in common. Each recurrent geodesic thus belongs to one and only one minimal set. The question arises as to how many different minimal sets there are on the given surface. That there are at least an enumerable infinity, follows at once from the fact that there are an enumerable infinity of closed geodesics. The number of minimal sets that do not consist simply of one closed geodesic still remains to be determined.

It has been seen that any geodesic lying wholly on S can be completely characterized by means of normal sets of sensed normal segments. It may be inquired whether or not any minimal set may not be characterized in terms of sensed normal segments, and if so, what is the explicit nature of the characterization. These questions seem to indicate the opening to an interesting field of inquiry.

HARVARD UNIVERSITY,
June, 1917

ON THE LOCATION OF THE ROOTS OF THE JACOBIAN OF TWO BINARY FORMS, AND OF THE DERIVATIVE OF A RATIONAL FUNCTION*

BY

J. L. WALSH

The present paper is an extension and in some respects a simplification of a recent paper published under the same title.† Both papers are based on a theorem (Theorem I, below) due to Professor Bôcher.‡ By means of the statical problem of determining the positions of equilibrium in a certain field of force, there are obtained some new results concerning the location of the roots of the jacobian of two binary forms relative to the location of the roots of the ground forms. Application is made to the roots of the derivative of a polynomial and to the roots of the derivative of a rational function. The present paper gives a proof and an application of a geometrical theorem (Theorem II) which may be not uninteresting.

Bôcher considers a number of fixed particles in a plane or by stereographic projection on the surface of a sphere, and supposes each particle to repel with a force equal to its mass (which may be positive or negative) divided by the distance. If the plane is taken as the Gauss plane, the following result is proved:§

THEOREM I. *The vanishing of the jacobian of two binary forms f_1 and f_2 of degrees p_1 and p_2 respectively determines the points of equilibrium in the field of force due to p_1 particles of mass p_2 situated at the roots of f_1, and p_2 particles of mass $-p_1$ situated at the roots of f_2.*

The jacobian vanishes not only at the points of no force, but also at the multiple roots of either form or a common root of the two forms; such a point is called a position of *pseudo-equilibrium.*

* Presented to the Society, Dec. 31, 1919.

† Walsh, these Transactions, vol. 19 (1918), pp. 291–298. This paper will be referred to as I.

‡ Maxime Bôcher, *A problem in statics and its relation to certain algebraic invariants,* Proceedings of the American Academy of Arts and Sciences, vol. 40 (1904), p. 469.

§ Bôcher's proof (l. c., p. 476) is reproduced in I, p. 291.

It is intuitively obvious that there can be no position of equilibrium very near any of the fixed particles, or very near and outside of a circle containing a number of fixed particles, all attracting or all repelling, if the other particles are sufficiently remote. We consider, then, a number of particles in a circle or more generally in a circular region. First we adjoin to the plane the point at infinity, and use the term *circle* to include point and straight line; then we define a *circular region* to be a closed region of the plane bounded by a circle, namely, the interior of a circle, the exterior of a circle including the point at infinity, a half plane, a point, or the entire plane. There will be no confusion in having the same notation for a circular region as for its boundary.

In the following development we shall use several lemmas.

LEMMA I. *The force at a point P due to k particles each of unit mass situated in a circular region C not containing P is equivalent to the force at P due to k coincident particles each of unit mass also in C.*

Denote by C' the inverse of C in the circle of unit radius and center P and by Q' the inverse of any point Q with regard to that circle. The force at P due to a particle at Q is in direction and magnitude PQ'. We replace k vectors PQ' by k coincident vectors having one terminal at P and the other at the center of gravity of the points Q'; these two sets of vectors have the same resultant. If any point Q is in the region C, its inverse Q' is in C', and the center of gravity of a number of such points Q' is also in C'. The inverse of this center of gravity is then in C.

LEMMA II. *In the field of force due to k positive particles at z_1, l positive particles at z_2, and $k+l$ negative particles at z_3, the only position of equilibrium is z_4 as determined by the cross-ratio*

$$\frac{(z_1 - z_2)(z_3 - z_4)}{(z_2 - z_3)(z_4 - z_1)} \equiv (z_1, z_2, z_3, z_4) = \frac{k+l}{k}.$$

The lemma is evidently true when one of the points z_1, z_2, z_3 is at infinity. The invariance of the positions of equilibrium under linear transformation follows from Theorem I and hence completes the proof.

We shall next prove a preliminary theorem, the proof of which is given in part by several succeeding lemmas.

THEOREM II. *If the envelopes of points z_1, z_2, z_3 are circular regions C_1, C_2, C_3 respectively, then the envelope of z_4, defined by the real constant cross-ratio*

$$\lambda = (z_1, z_2, z_3, z_4)$$

*is also a circular region.**

* The term *envelope* is used to denote the set of points which is the totality of positions assumed by each of the points z_1, z_2, z_3, z_4; the points z_1, z_2, z_3 are supposed to vary independently.

The proof of Theorem II which is presented in detail has some advantages and some dis-

We denote the envelope of z_4 by C_4, and we must show that C_4 is a region bounded by a single circle. First we consider several special cases of the theorem. If C_1, C_2, and C_3 are distinct points, C_4 is a point. If any of the regions C_1, C_2, C_3 is the entire plane, C_4 is also the entire plane. If $\lambda = 0$ and if C_1 and C_2 have a point in common, C_4 is the entire plane. If $\lambda = 0$ and C_1 and C_2 have no point in common, $z_3 = z_4$ and so C_4 coincides with C_3. If $\lambda = \infty$ and C_2 and C_3 have a common point, C_4 is the entire plane. If $\lambda = \infty$ and C_2 and C_3 have no common point, C_4 and C_1 are identical. If $\lambda = 1$ and C_1 and C_3 have a common point, C_4 is the entire plane. If $\lambda = 1$ yet C_1 and C_3 have no common point, C_4 is identical with C_2. In the sequel, unless it is explicitly stated to the contrary, we suppose λ to have none of the values $0, 1, \infty$. It follows that no two of the points z_1, z_2, z_3, z_4 coincide unless three of them coincide.

Except in the trivial case that C_1, C_2, C_3 are points, C_4 is evidently a two-dimensional continuum and is not necessarily the entire plane. The envelope C_4 is connected, for to join any pair of points z_4', z_4'' in C_4 by a curve in C_4, we need merely to choose any set of points corresponding to each, z_1', z_2', z_3'; z_1'', z_2'', z_3'', in the proper regions. Join z_1' and z_1'' by a continuous curve which lies in C_1, and similarly join z_2' and z_2'', and z_3' and z_3'', by continuous curves in C_2 and C_3 respectively. Allow z_1, z_2, z_3 to move from z_1', z_2', z_3' to z_1'', z_2'', z_3'' along these respective curves. The point z_4 corresponding moves from z_4' to z_4'' in C_4 and along a curve which is continuous because z_4 is a linear function of z_1, z_2, z_3.

Our next remark is stated explicitly as a lemma. It is readily stated and established for regions whose boundaries are curves much more general than circles, but we consider here merely the form under the hypothesis of Theorem II and for application to the proof of that theorem.

---

advantages over the following suggested method of proof. The theorem is evidently true when C_1, C_2, and C_3 are points. The theorem is easily proved when C_1 and C_2 are points but C_3 is not a point. By taking the envelope of the circular region C_4 in the preceding degenerate case, the theorem can be proved when C_1 is a point but neither C_2 nor C_3 is a point. The envelope of the region C_4 in this last degenerate case, as z_1 is allowed to vary over a region C_1 not a point, gives the envelope of z_4 for the theorem in its generality. I have not been able to carry through the actual analytic determination of the envelope by this method because the algebraic work is too laborious.

This suggested method of proof, however, shows at once that the boundary of the region C_4 in the general case is an algebraic curve or at least part of an algebraic curve.

It seems to me likely that Theorem II is true also when λ is imaginary, but I have not carried through the proof in detail.

In general the relation of the regions C_1, C_2, C_3, C_4 is not reciprocal. For example if C_1 is a point but neither C_2 nor C_3 is a point and if these regions lead to the fourth region C_4, then if we choose the circular regions C_2, C_3, C_4 as the original circular regions of the lemma, we cannot for any choice of λ be led to the region C_1. This lack of reciprocality does not depend on the degeneracy of one of the regions C_1, C_2, C_3, C_4.

Lemma III. *If the point* z_4 *is on but not at a vertex of the boundary of* C_4,* *then any set of points* z_1, z_2, z_3 *corresponding lie on the boundaries of the respective regions* C_1, C_2, C_3; *the circle* C *through the points* z_1, z_2, z_3, z_4 *cuts the circles* C_1, C_2, C_3 *all at angles of the same magnitude; and if* C *is transformed into a straight line, the lines tangent to the circles* C_1, C_2, *and* C_3 *at the points* z_1, z_2, z_3 *respectively are parallel.*

The following proof is formulated only for the general case that none of the circles C_1, C_2, C_3 is a null circle, but no essential modification is necessary to include the degenerate cases.

When z_2 and z_3, and also the circle C are kept fixed, a continuous motion of z_1 along C also causes z_4 to move continuously along C. If the direction of motion of z_1 is reversed, the direction of motion of z_4 is also reversed. Hence z_4 is not on the boundary of C_4 unless z_1 is on the boundary of C_1, and as can be shown in an analogous manner, not unless z_2 and z_3 are on the boundaries of C_2 and C_3 respectively. The region C_4 is closed since the regions C_1, C_2, and C_3 are closed.

Let P be any point of the boundary of C_4. Transform P to infinity, so that the corresponding points z_1, z_2, z_3 lie on the same line L. We assume at first that L is not tangent to any of the circles C_1, C_2, C_3 nor to the boundary of C_4. The relative positions of the points z_1, z_2, z_3 on L together with the sense along L in which the region C_1 extends from z_1 determine uniquely the sense along L in which the regions C_2, C_3, C_4 must extend from z_2, z_3, P respectively. There is evidently a segment of L terminated by P composed entirely of points in C_4. If the lines tangent to the circles C_2 and C_3 at the points z_2 and z_3 are not parallel, it is possible slightly to rotate L about z_1 in one direction or the other into a new position L' and to determine a point z_2'' on L' and on the circle C_2 and a point z_3'' on L' and *interior to the region* C_3 such that the triangles $z_1 z_2 z_2''$ and $z_1 z_3 z_3''$ are similar and hence we have the relation

$$(z_1, z_2'', z_3'', P) = \lambda.$$

Then z_3'' can be moved in either sense along the line L' and still remain in its proper envelope, so there are corresponding points z_4'' on L' in either sense from P. Moreover, this is true for every position of L' if the angle from L to L' is in the proper sense and is sufficiently small, so if we transform P to the finite part of the plane and z_1 to infinity and notice that the lines L' are lines through the point P, it becomes evident that there are points z_4 in the neighborhood of P on any line L' through P which lies within a certain sector whose vertex is P, and there are points z_4 on L' in both directions

* It is of course true that the boundary of C_4 has no vertices, but that fact has not yet been proved.

from P. Hence if P is actually on the boundary of C_4, it must lie at a vertex of that boundary.*

The proof thus far has been formulated to prove that when P is at infinity the lines tangent to the circles C_2 and C_3 at z_2 and z_3 are parallel. The notation of the proof can easily be modified to show that the lines tangent to the circles C_1 and C_2 at z_1 and z_2 are parallel, and hence the lines tangent to C_1, C_2, C_3 at z_1, z_2, z_3 are parallel.

This same method of reasoning is readily used to prove that if the circle C of the lemma is tangent to one or two of the circles C_1, C_2, C_3 at the respective points z_1, z_2, z_3 but is not tangent to all these circles, the boundary of C_4 has a vertex at z_4. The circle C is not tangent to the boundary of C_4 unless C is tangent to C_1, C_2, and C_3. This consideration completes the proof of Lemma III.

It is desirable to make a revision in our use of the term *angle between two circles.* With Coolidge,† we consider circles to be described by a point moving in a counter-clockwise sense, and define the angle between two circles to be the angle between the half-tangents drawn at the intersection in the sense of description of the circles. When we are concerned with a single straight line, either sense may be given to it. We shall use this convention in proving the following lemma, which is a result purely of circle geometry which has not necessarily any connection with Theorem II. As stated and proved, it is slightly more general than is necessary for its application in the proof of that theorem.

LEMMA IV. *Suppose a variable circle C either to cut three distinct fixed non-coaxial circles C_1, C_2, C_3 all at the same angle or to cut a definite one of those circles at an angle supplementary to the angle cut on the other two. If the points z_1, z_2, z_3 are chosen as intersections of C with C_1, C_2, C_3 respectively such that when C is transformed into a straight line the lines tangent to C_1, C_2, C_3 at z_1, z_2, z_3 are all parallel, then the locus of the point z_4 defined by the real constant cross-ratio*

$$\lambda = (z_1, z_2, z_3, z_4)$$

is a circle C_4 which is also cut by C at an angle equal or supplementary to the angles cut on C_1, C_2, C_3.‡

This lemma is not true if the circles C_1, C_2, C_3 are coaxial circles having no point in common. For transform these circles into concentric circles. Then

* The method of proof used in this paragraph was suggested to me by Professor Birkhoff.

† *A treatise on the circle and the sphere*, p. 108.

‡ We remark that the circle C_4 can be constructed by ruler and compass whenever λ is rational or in fact whenever λ is given geometrically. For the circle C can be constructed by ruler and compass in any position; cf. Coolidge, l. c., p. 173. Hence we can determine any number of sets of points z_1, z_2, z_3 and therefore construct any number of points z_4, which enables us to construct C_4.

the circle C is a straight line orthogonal to these circles, C has two intersections with each, and on any particular circle C the points z_1, z_2, z_3 may be chosen on their proper circles so as to lead to four circles of type C_4, in general distinct, and concentric with C_1, C_2, C_3. All these four circles of type C_4 form the locus of points z_4. The situation is essentially the same if C_1, C_2, C_3 are coaxial circles having two common points; we are led to four circles C_4 which are in general distinct. But if we suppose C to vary continuously and also the points z_1, z_2, z_3, z_4 each to vary in one sense continuously, although of course we allow these points to go to infinity but not to occupy any position more than once, the lemma is true even for coaxial circles having no point or two points in common. These situations are included in the detailed treatments given under Cases I and II below.

This lemma breaks down also if the circles C_1, C_2, C_3 are coaxial circles all tangent at a single point, for we can consider the three points z_1, z_2, z_3 to coincide at that point; any circle C through that point satisfies the conditions of the lemma, any point of C can be chosen as z_4, whence it appears that the locus of z_4 is then the entire plane. But if we make not only our previous convention but in addition the convention that not all of the points z_1, z_2, z_3 shall lie at a point common to the three circles unless the fourth point coincides with them, then the lemma remains true. This situation is treated in detail under Case IV below.

The lemma is true but trivial in the degenerate cases $\lambda = 0, 1$, or ∞, for in these cases z_4 coincides with z_3, z_2, or z_1 respectively. The case that C_1, C_2, and C_3 are all null circles is likewise trivial. In the consideration of other cases we shall use the following theorem:

THEOREM. *If three circles be given not all tangent at one point, the circles cutting them at equal angles form a coaxial system, as do those cutting one at angles supplementary to the angles cut on the other two.**

Then as the circle C of Lemma IV varies, it always belongs to a definite coaxial system, unless C_1, C_2, C_3 are all tangent at a single point. This system may consist of (Case I) circles through two points, (Case II) non-intersecting circles, or (Case III) circles tangent to a line at a single point. Under Case IV will be treated the situation when C_1, C_2, C_3 are all tangent at a point. We consider these cases in order.

In Case I, transform to infinity one of the two points through which the coaxial family C passes, so that this family becomes the straight lines through a finite point q of the plane. In general q will be a center of similitude for each pair of the circles C_1, C_2, and C_3. These circles may or may not surround q.

* This statement differs from that of Coolidge, l. c., p. 111, Theorem 219, for we have adjoined to the plane the point at infinity. Theorem 220 seems to be erroneous; compare the four circles C_1, C_2, C_3, C_4 of Lemma IV.

Let z_4 be any point corresponding to the points z_1, z_2, z_3 on C_1, C_2, C_3 respectively. These four points lie on the line qz_4, and we have supposed that the lines tangent to C_1, C_2, C_3 at the points z_1, z_2, z_3 are parallel. Then when the line qz_4 (that is, the circle C) rotates about q, it will be seen that the point z_4 as determined by its constant cross-ratio with z_1, z_2, z_3 will trace a circle C_4 such that q is a center of similitude for any of the pairs of circles C_1, C_2, C_3, C_4. If these circles do not surround q, they have two common tangents belonging to the family C, and the properly chosen cross-ratio of the points of tangency is λ. If C_1, C_2, and C_3 are coaxial, C_4 is coaxial with them. Perhaps it is worth noticing that any circle C_4 such that q is a center of similitude for any pair of the circles C_1, C_2, C_3, C_4 is the circle C_4 of the lemma for a proper choice of λ; in particular C_4 may be the point q or the point at infinity.

Under Case I there are some special situations to be included. If one or more of the circles C_1, C_2, C_3 passes through q, then each of the other circles if not a null circle either is tangent to that circle at q or is a line parallel to the line tangent to that circle at q. If two of the original circles, for definiteness C_1 and C_2, are tangent at q and the other circle C_3 is a line parallel to their common tangent at q, then either z_4 coincides with z_1 and z_2 at q, or z_3 remains at infinity during the motion of C while z_4 traces a circle coaxial with C_1 and C_2; in particular this circle C_4 may be the null circle q. The four circles C_1, C_2, C_3, C_4 have a common tangent circle, namely the line tangent to C_1, C_2, C_4 at q. In the case just considered, one of the circles which passes through q, for definiteness C_1, may be tangent at q to the second circle C_2 which is a straight line. The circle C_3 is a line parallel to C_2. When the circle C varies, z_4 coincides with z_1 and z_2 at q, z_4 coincides with z_2 and z_3 at infinity, or the circle C coincides with C_2, z_1 with q, and z_3 with the point at infinity, while z_2 traces the line C_2 and hence z_4 also traces C_2. The circles C_1, C_2, C_3, C_4 have a common tangent circle C_2. If one of the original circles, for definiteness C_1, passes through q and the circles C_2 and C_3 are lines parallel to the tangent to C_1 and q, then the circle C_4 is a circle coaxial with C_2 and C_3 which may be the point at infinity. The four circles C_1, C_2, C_3, C_4 have as common tangent circle the line tangent to C_1 at q.

The general situation of Case I is not essentially changed and requires no further discussion if one of the circles C_1, C_2, C_3 is a point (q or the point at infinity) or if two of them are points (q and the point at infinity), except when at least one of the null circles lies on one of the non-null circles. In particular, if two circles, for example C_1 and C_2, are null circles and one of them (say C_2) lies on the non-null circle C_3, the locus of z_4 is a circle C_4 tangent to the circle C_3 at the point C_2. If the two null circles C_1 and C_2 both lie on the non-null circle C_3, the circle C is effectually the circle C_3, and C_4 coincides with C_3.

The special situations which we have considered under Case I may similarly degenerate by having one of the original circles a null circle. We shall discuss merely some typical examples. If C_1 and C_2 are tangent at q and C_3 is a null circle at infinity, C_4 is a circle tangent to C_1 and C_2 at q and may be the point q itself. If C_1 is a null circle at q, if C_2 is a circle passing through q, and if C_3 is a line parallel to the tangent to C_2 at q, C_4 is a circle tangent to C_2 at q. If C_1 is a null circle at q, if C_2 is a line passing through q, and C_3 is a line parallel to C_2, then C is essentially the single circle C_2, and C_4 coincides with C_2.

In Case II, the coaxial family C is composed of circles having no point in common, and hence there are two null circles of the family. Transform one of these null circles to infinity, so that the family C becomes a family of circles with a common center p. In the general case, the circles C_1, C_2, and C_3 are all of equal radii and any of them can be brought into coincidence with any other of them by a rotation about p. The point p is outside, on, or within all three circles according as it is outside, on, or within any one of them. Choose any point z_4 of the lemma; then z_1, z_2, z_3, z_4 lie on the circle C whose center is p. As C varies, its radius simply increases or decreases, and z_1, z_2, z_3 rotate about p so that the angles z_2pz_3, z_3pz_1, z_1pz_2 remain constant. Hence z_4 traces a circle C_4 whose radius is equal to the common radius of C_1, C_2, and C_3; moreover any two of the four circles C_1, C_2, C_3, C_4 can be brought into coincidence by a rotation about p. The four circles have two common tangent circles which belong to the family C, one of which may be the point p. The properly chosen cross-ratio of the points of tangency of a tangent circle is λ. Any circle is the circle C_4 of the lemma for a proper choice of λ provided it can be brought into coincidence with any of the circles C_1, C_2, C_3 by a rotation about p.

Another situation that may arise under Case II is that C_1, C_2, and C_3 are straight lines (that is, coaxial circles) through p and the point at infinity; then the locus of z_4 is a circle C_4 coaxial with them. There remains also the possibility that C_1, C_2, C_3 are straight lines all at the same distance from p. Then the circle C_4 is a line also at this same distance from p. There is a circle belonging to the family C which is tangent to C_1, C_2, C_3, C_4, and as before the cross-ratio of the points of contact is λ.

In Case III, the circles C belong to a coaxial family of circles all tangent at a point n, which point we transform to infinity. The circles C become parallel lines and in general C_1, C_2, C_3 become equal circles whose centers are collinear. As C moves parallel to itself, the points z_1, z_2, z_3 remain at equal distances from each other. The locus of z_4 either is a circle C_4 equal to C_1, C_2, and C_3 whose center is collinear with their centers or is the point at infinity. The four circles have two common tangent circles which belong

to the family C, and the cross-ratio of the points of tangency of each of these circles is λ.

A degenerate case that should be mentioned is that the point n itself is one of the circles C_1, C_2, C_3. The results are essentially the same as in the general situation. In both the degenerate and the general situations any circle C_4 equal to C_1, C_2, C_3 and whose center is collinear with their centers is the circle C_4 of the lemma if λ is properly chosen.

A special case also occurs if one of the original circles, for definiteness C_1, is a straight line and the other two circles are straight lines parallel to the reflection of C_1 in any of the circles C. When C varies, either z_4 coincides with z_2 and z_3 at infinity, or z_1 is at infinity and z_4 traces a line parallel to C_2 and C_3.

A degenerate case occurs if one of the original circles, say C_3, is the point at infinity, while C_1 and C_2 are the reflections of each other in one of the circles C. Under the conditions of the lemma z_4 must coincide with z_3 at infinity, so C_4 coincides with C_3.

In Case IV, the circles C_1, C_2, C_3 are all tangent at a point m. Transform m to infinity, so that in any non-degenerate case C_1, C_2, C_3 become parallel lines. Under our convention that not all of the points z_1, z_2, z_3 shall lie at m unless z_4 coincides with them, we are led to four circles (in general distinct) according as we allow any one of the points z_1, z_2, z_3 or none of them constantly to lie at infinity. The additional convention already made that z_1, z_2, z_3, z_4 shall vary continuously in one sense and never coincide with any previous position enables us to choose simply one of these circles. The circle C is any straight line, and z_4 is either the intersection of C with a straight line C_4 parallel to C_1, C_2, C_3 or if none of the points z_1, z_2, z_3 is at infinity, z_4 may be constantly the point at infinity. The circles C_1, C_2, C_3, C_4 are all tangent at m.

Under Case IV should be mentioned the degenerate case that one of the circles C_1, C_2, C_3 is a null circle lying at the point of tangency of the other two circles. Our conventions enable us to choose a circle C_4 coaxial with C_1, C_2, C_3.

The proof of Lemma IV is now complete. It will be noticed that except in the special and degenerate cases, the result is entirely symmetric with respect to the four circles C_1, C_2, C_3, C_4. If we commence by choosing any three of those four circles and choose λ properly we shall be led to the other circle. If the last clause in the statement of the lemma is omitted, the lemma is true even if λ is not real.

There is a lemma corresponding to Lemma IV if we suppose two of the original circles, for example C_1 and C_2, to coincide, but suppose C_3 not to coincide with them. If we leave aside the easily treated cases $\lambda = 0, 1$, or ∞,

we find either that the points z_1 and z_2 coincide on C_1, in which case z_4 coincides with them and traces the circle C_1, or that if C_1 is a non-null circle z_1 and z_2 do not coincide. In the latter case we are supposing the tangents to C at z_1 and z_2 to be parallel if C is transformed into a straight line and hence C must be orthogonal to C_1 and therefore by the conditions of the lemma also orthogonal to C_3. As before, when the circle C varies it constantly belongs to a definite coaxial system. The reader will easily treat the cases corresponding to Cases I, II, and III above, and also the degenerate case that C_3 is a null circle lying on C_1 and C_2. The results in the general case are quite analogous to the previous results if we notice that C_1, C_2, and C_3 are coaxial. For if C_3 is not a null circle, C cuts C_3 in two distinct points, and by their cross-ratio with z_1 and z_2 these lead to *two distinct circles* C_4 in addition to the circle C_1. Both of these new circles C_4 belong to the coaxial family determined by C_1 and C_3; as C moves it is constantly orthogonal to C_4 as well as to C_1, C_2, C_3. In general, then, the locus of z_4 when C_1 and C_2 coincide is C_1 and two other circles of the coaxial family determined by C_1 and C_3. These two other circles may in a degenerate case coincide, as the reader can easily determine. The convention formerly made, that the points z_1, z_2, z_3, z_4 vary in one sense continuously will of course restrict the locus of z_4 simply to one circle.

When the three circles C_1, C_2, C_3 coincide, we must consider C to coincide with them, or else at least two of the points z_1, z_2, z_3 to coincide and hence z_4 to coincide with them. That is, the circle C_4 corresponding to the circle C_4 of the lemma is the circle C_1.

Lemmas III and IV with the discussion supplementary to the latter do not give us immediately all the material necessary for the proof of Theorem II. For if C_1, C_2, C_3 are coaxial there are four circles, not necessarily all distinct, of the type C_4 of the lemma. If C_1, C_2, C_3 are not coaxial there are also four circles, not necessarily all distinct, of the type C_4 of the lemma, according as C cuts all the circles C_1, C_2, C_3 at equal angles or cuts one at an angle supplementary to the angle cut on the other two. It is conceivable that the boundary of the region C_4 of Theorem II should consist of arcs of more than one distinct circle; we proceed to show that this is in fact never the case.* The following lemma is essential in our proof.

* Whether the boundary of the region C_4 corresponds to motion of C cutting the three original circles at the same angle or a definite one of those circles at an angle supplementary to the angle cut on the other two depends on the relative positions of those circles, on whether the various regions are interior or exterior to their bounding circles, and on the value of λ—in short, on the order of the points z_1, z_2, z_3, z_4 on the circle C. When the regions C_1, C_2, C_3 are mutually external it is easy to prove by reasoning similar to that used in the proof of Lemma III that an arc of only one of the circles of type C_4 can be a part of the boundary of the region C_4. This fact can also be proved in the general case by that same method of reasoning, but the proof given in detail below is perhaps more satisfactory. It is desirable

LEMMA V. *In Theorem II, whenever the envelope of z_4 is not the entire plane, there is a circle S orthogonal to the four circles C_1, C_2, C_3, C_4.*

Whenever the regions C_1, C_2, C_3 have a common point, we may consider z_1, z_2, z_3 to coincide at that point, and consider the cross-ratio of any point z_4 in the plane with those three points to have the value λ, so the envelope of z_4 is the entire plane. In any other case there is a circle S orthogonal to the circles C_1, C_2, C_3. If not every pair of these three original circles intersect, choose two of them which do not intersect, and there will be two points inverse respecting both circles (these points are the null circles of the coaxial family determined by the two circles). Take the inverse of one of those points in the third of the original circles and pass a new circle S through all three points. Then S is orthogonal to the three original circles. If each of the circles C_1, C_2, C_3 has a point in common with the other two, we can transform two of the circles into straight lines (if one of the circles is a null circle the other two circles pass through that null circle and hence the region C_4 is the entire plane). If these two lines are not parallel, the third circle cannot be a straight line nor can it surround the intersection of the other two lines. Hence there is a circle orthogonal to all three circles. If the two lines are parallel the third circle cannot be a straight line. Then there is a circle, in this case a straight line, orthogonal to all three circles. This completes the proof of Lemma V.

Let us transform into a straight line any particular circle S orthogonal to the three original circles and let us suppose not every point of S to be a point of the region C_4; for definiteness assume the point at infinity not to belong to C_4. The positions which each of the three points z_1, z_2, z_3 of Theorem II may occupy fill an entire segment of S, and hence the points z_4 on S which correspond to points z_1, z_2, z_3 on S fill an entire segment of S; we denote this segment by σ. The terminal points of the segment σ are the intersections of S with one of the circles of type C_4 of Lemma IV; we denote that circle by C_4' and the other three circles of that type by C_4'', C_4''', C_4''''. The entire configuration is symmetric with respect to S, so the centers of all the circles $C_4', C_4'', C_4''', C_4''''$ lie on S. Moreover, S belongs to all four types of circles C of Lemma IV, since it is orthogonal to C_1, C_2, C_3. Hence the intersections of all the circles C_4'', C_4''', C_4'''' are points z_4 which correspond to points z_1, z_2, z_3 lying on S, and hence all those intersections lie on the segment σ. Then of the circles C_4'', C_4''', C_4'''' each is interior to or coincident with C_4'.

Either the entire interior or the entire exterior of each of the circles C_4', C_4'', C_4''', C_4'''' belongs to the region C_4. For the points z_4 which correspond to

that most of the material making up that proof should be given anyway, as a test whether the region C_4 is the entire plane, as giving a ruler-and-compass construction for the circle C_4, and as describing more in detail the entire situation with which we are concerned.

points z_1, z_2, z_3 in the proper regions and on the circle C of Lemma IV fill an entire arc of C, extending from one intersection of C with the circle C_4 to the other intersection. The entire exterior of our circle C_4' does not belong to the region C_4, for the point at infinity does not belong to that region. Hence the entire interior of C_4' does belong to the region C_4. No point external to C_4' can be a point of the boundary of C_4, for none of the circles C_4'', C_4''', C_4'''' has a point exterior to C_4'. Hence the region C_4 is the interior of C_4', under our assumption that not every point of S belongs to the region C_4.

Let us notice that we can allow any or all of the circles C_1, C_2, C_3 to move continuously so as to remain orthogonal to S, so as never to intersect any former position, and so as always to enlarge the regions C_1, C_2, C_3. Then the circle C_4' grows larger and larger, never intersecting its former position, until it becomes the point at infinity, in which case the region C_4 is the entire plane. If the regions C_1, C_2, C_3 are enlarged still further, the region C_4 still remains the entire plane.

Whether or not we assume that not every point of S belongs to the region C_4, we can start with a situation in which not every point of S is a point of C_4 and enlarge the regions C_1, C_2, C_3 in the manner described so as to attain any situation desired in which the region C_4 is not the entire plane. At every stage the region C_4 is a circular region. This completes the proof of Theorem II. We have also obtained a test whether or not the region C_4 is the entire plane. *A necessary and sufficient condition that the region C_4 of Theorem II be the entire plane is that the point z_4 may occupy any position on S and still correspond to points z_1, z_2, z_3 in their proper envelopes and also on S.*

The preceding developments give a comparatively simple ruler-and-compass construction for the circle C_4, whenever λ is rational or is given geometrically. The circle S can be constructed by ruler and compass.* The two points of intersection of S and C_4 can be determined by means of their cross-ratio with properly chosen intersections of S and C_1, C_2, C_3. Since S and C_4 are orthogonal, C_4 can then be constructed.

We shall apply Theorem II in proving our principal theorem.

THEOREM III. *Let f_1 and f_2 be binary forms of degrees p_1 and p_2 respectively, and let the circular regions C_1, C_2, C_3 be the respective envelopes of m roots of f_1, the remaining $p_1 - m$ roots of f_1, and all the roots of f_2. Denote by C_4 the circular region which is the envelope of points z_4 such that*

$$(z_1, z_2, z_3, z_4) = \frac{p_1}{m},$$

when z_1, z_2, z_3 have the respective envelopes C_1, C_2, C_3. Then the envelope of

* Coolidge, l. c., p. 173.

the roots of the jacobian of f_1 *and* f_2 *is the region* C_4, *together with the regions* C_1, C_2, C_3 *except that among the latter the corresponding region is to be omitted if any of the numbers* m, $p_1 - m$, p_2 *is unity. If a region* C_i $(i = 1, 2, 3, 4)$ *has no point in common with any other of those regions which is a part of the envelope of the roots of the jacobian, it contains of those roots precisely* $m - 1$, $p_1 - m - 1$, $p_2 - 1$, *or* 1 *according as* $i = 1, 2, 3$, *or* 4.

We shall first show by the aid of Lemmas I and II and of Theorems I and II that no point not in C_1, C_2, C_3, or C_4 can be a root of the jacobian. For if a point z_4 is not in C_1, C_2, or C_3 and is a root of the jacobian, it is a position of equilibrium and not of pseudo-equilibrium. The force at z_4 will not be changed if we replace the particles in each of the regions C_1, C_2, C_3 by the same number of coincident particles at points z_1, z_2, z_3 in the respective regions. Then z_4 is a position of equilibrium in the new field of force and hence by Lemma II we have

$$(z_1, z_2, z_3, z_4) = \frac{p_1}{m},$$

and therefore z_4 lies in C_4.

Any point in C_4 can be a root of the jacobian, for we need merely find points z_1, z_2, z_3 in the regions C_1, C_2, C_3 such that

$$(z_1, z_2, z_3, z_4) = \frac{p_1}{m}$$

and allow all the roots of the ground forms in each of those regions to coincide at those points. Any point of a region C_1, C_2, C_3 which is the envelope of more than one root of a ground form can be a position of pseudo-equilibrium and hence a root of the jacobian. If any of the regions C_1, C_2, C_3 is the envelope of merely one root of a ground form, then no point in that region but not in any other of the regions C_1, C_2, C_3, C_4 can be a position of equilibrium or of pseudo-equilibrium and hence no such point can be a root of the jacobian. If a point is common to two of the regions C_1, C_2, C_3, C_4 it is a point of C_4 and hence is a point of the envelope of the roots of the jacobian.

We have now proved the theorem except for its last sentence, to the demonstration of which we now proceed. When the roots of the ground forms in the regions C_1, C_2, C_3 coincide, the regions C_1, C_2, C_3, C_4 contain respectively the following numbers of roots of the jacobian: $m - 1$, $p_1 - m - 1$, $p_2 - 1$, 1. The roots of the jacobian vary continuously when the roots of the ground forms vary continuously; no root of the jacobian can enter or leave any of the regions C_1, C_2, C_3, C_4 which has no point in common with any other of those regions which is a part of the envelope of the roots of the jacobian.

The proof of Theorem III is now complete.* It applies to the sphere as well as the plane, since everything essential in the theorem is invariant under stereographic projection.

Instead of considering primarily the jacobian of two binary forms as heretofore, we may consider a rational function $f(z)$, introduce homogeneous coördinates, and compute the value of the derivative $f'(z)$ in terms of J, the jacobian of the binary forms which are the numerator and denominator of $f(z)$. We find that *the roots of $f'(z)$ are the roots of J and a double root at infinity, except that when one of these points is also a pole of $f(z)$ it cannot be a root of $f'(z)$.*† Application of Theorem III gives a theorem analogous to Theorem III, but which we state in a form slightly different from the statement of that theorem.

THEOREM. *If the circular regions C_1, C_2, C_3 contain respectively m roots (or poles) of a rational function $f(z)$ of degree p, all the remaining roots (or poles) of $f(z)$, and all the poles (or roots) of $f(z)$, then all the roots of $f'(z)$ lie in the regions C_1, C_2, C_3, and a fourth circular region C_4 determined as the envelope of points z_4 such that*

$$(z_1, z_2, z_3, z_4) = \frac{p}{m},$$

while the envelopes of z_1, z_2, z_3 are respectively C_1, C_2, C_3,—except that there are two roots at infinity if $f(z)$ has no pole there. Except for these two additional roots, if any of the regions C_i $(i = 1, 2, 3, 4)$ has no point in common with any other of those regions which contains a root of $f'(z)$, then that region contains the following number of roots of $f'(z)$ for $i = 1, 2, 3, 4$ respectively:

$$m - 1, \qquad p - m - 1, \qquad q_3 - 1, \qquad 1;$$

or

$$q_1 - 1, \qquad q_2 - 1, \qquad p - 1, \qquad 1,$$

according as C_1 contains m roots or m poles of $f(z)$; here q_i indicates the number of distinct poles of $f(z)$ in C_i.

Perhaps the following special cases of this theorem are worth stating explicitly.

If $f(z)$ is a rational function whose m_1 finite roots (or poles) lie on or within a circle C_1 with center α_1 and radius r_1 and whose m_2 finite poles (or roots) lie on or within a circle C_2 with center α_2 and radius r_2, and if $m_1 > m_2 > 0$, then

* It may be noticed that this proof does not explicitly use the fact that C_4 is a circular region.

If C_1, C_2, C_3 are coaxial circles with no point in common, Theorem III reduces essentially to Theorem II (I, p. 294). If $m = 0$ or $p_1 - m = 0$, the regions C_1, C_2, and C_4 can be considered to coincide; this gives Theorem III (I, p. 296), which is due to Bôcher.

† See I, p. 297.

all the finite roots of $f'(z)$ lie in C_1, C_2, and a third circle C_3 whose center is

$$\frac{m_1 \alpha_2 - m_2 \alpha_1}{m_1 - m_2}$$

and radius

$$\frac{m_1 r_2 + m_2 r_1}{m_1 - m_2}.$$

*If $f(z)$ has no finite multiple poles, and if C_1, C_2, C_3 are mutually external, they contain respectively the following numbers of roots of $f'(z)$: $m_1 - 1$, $m_2 - 1$, 1. Under the given hypothesis, if $m_1 = m_2$ and if C_1 and C_2 are mutually external, these circles contain all the finite roots of $f'(z)$.**

If $f(z)$ is a polynomial m_1 of whose roots lie on or within a circle C_1 whose center is α_1 and radius r_1, and if the remaining m_2 roots lie on or within a circle C_2 whose center is α_2 and radius r_2, then all the roots of $f'(z)$ lie on or within C_1, C_2, and a third circle C_3 whose center is

$$\frac{m_1 \alpha_2 + m_2 \alpha_1}{m_1 + m_2}$$

and radius

$$\frac{m_1 r_2 + m_2 r_1}{m_1 + m_2}.$$

If these circles are mutually external, they contain respectively the following number of roots of $f'(z)$: $m_1 - 1$, $m_2 - 1$, 1.

If $f(z)$ is a polynomial of degree n with a k-fold root at P, and with the remaining $n - k$ roots in a circular region C, then all the roots of $f'(z)$ lie at P, in C, and in a circular region C' obtained by shrinking C toward P as center of similitude in the ratio $1 : k/n$. If C and C' have no point in common they contain respectively $n - k - 1$ roots and 1 root of $f'(z)$.†

A special case of this last theorem is the following

THEOREM. *If a circle includes all the roots of a polynomial $f(z)$, it also includes all the roots of $f'(z)$.*

* A more restricted theorem than this has been proved not merely for rational functions but also for the quotient of two entire functions. See M. B. Porter, *Proceedings of the National Academy of Sciences*, vol. 2 (1916), pp. 247, 335.

There is no theorem analogous to the theorem of the present paper if $m_1 = m_2$ and if C_1 and C_2 are not mutually external. For we may consider all the roots and all the poles of $f(z)$ to coincide, so that $f(z)$ reduces to a constant and every point of the plane is a root of $f'(z)$.

† This theorem is true whether the circle C surrounds, passes through, or does not surround P, and whether the region C is interior or exterior to the circle C. The special case where P is the center of the circle C and the region C is external to that circle was pointed out in a footnote, I, p. 298. The special case where C does not surround P and the region C is interior to the circle C was pointed out to me by Professor D. R. Curtiss.

The latter theorem is equivalent to the well-known theorem of Lucas:

If all the roots of a polynomial $f(z)$ lie on or within any convex polygon, then all the roots of $f'(z)$ lie on or within that polygon.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.,
May, 1920.

ON FUNCTIONS OF CLOSEST APPROXIMATION*

BY

DUNHAM JACKSON

1. **Introduction.** The determination of the polynomial of specified degree which gives the best approximation to a given continuous function $f(x)$ in a given interval (a, b) depends on the meaning attached to the phrase "best approximation." The polynomial for which the maximum of the absolute value of the error is as small as possible is known as the Tchebychef polynomial corresponding to $f(x)$, and has been extensively studied.† The polynomial which reduces the integral of the square of the error to a minimum is obtained by taking the sum of the first terms in the development of $f(x)$ in Legendre's series,‡ and its properties are of course also well known.

The following pages are devoted to a study of the polynomial for which the integral of the mth power of the error is a minimum, where m is any even positive integer, or, more generally, the integral of the mth power of the absolute value of the error, where m is any real number greater than 1. It is found that some of the familiar properties of the approximating function in the case $m = 2$ are carried over to the other values of m. It is shown further, and this is the principal conclusion of the paper, that the polynomial of approximation corresponding to the exponent m approaches the Tchebychef polynomial as a limit when m becomes infinite. The discussion is put in such a form as to apply also to approximation by finite trigonometric sums,§ or more generally to approximate representation by linear combinations of an arbitrary set of linearly independent continuous functions, having such further proper-

* Presented to the Society, April 10, 1920.

† Cf., e.g., Kirchberger, *Ueber Tchebychefsche Annäherungsmethoden*, Dissertation, Göttingen, 1902; Borel, *Leçons sur les fonctions de variables réelles et les développements en séries de polynomes*, pp. 82–92.

‡ Cf., e.g., Gram, *Ueber die Entwickelung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate*, Journal für die reine und angewandte Mathematik, vol. 94 (1883), pp. 41–73.

§ For the extension of Tchebychef's theory to the case of trigonometric approximation, see, e.g., Fréchet, *Sur l'approximation des fonctions par des suites trigonométriques limitées*, Comptes Rendus, vol. 144 (1907), pp. 124–125; J. W. Young, *General theory of approximation by functions involving a given number of arbitrary parameters*, these Transactions, vol. 8 (1907), pp. 331–344; Fréchet, *Sur l'approximation des fonctions continues périodiques par les sommes trigonométriques limitées*, Annales de l'École Normale Supérieure, ser. 3, vol. 25 (1908), pp. 43–56.

ties, in the case of the final theorem, as to insure the uniqueness of the best approximating function in the sense of Tchebychef.* It will be apparent that even this general treatment can be extended in various directions, of which nothing more will be said here. The force of the conclusions will be most readily appreciated, on the other hand, if they are made specific by identifying the functions $p_1(x), p_2(x), \cdots, p_n(x)$ of the text with the quantities $1, x, \cdots, x^{n-1}$, and $\phi(x)$ with an arbitrary polynomial of degree $n-1$.

2. **First lemma on bounds of coefficients.** Let

$$p_1(x), p_2(x), \cdots, p_n(x)$$

be n functions of x, continuou. throughout the interval

$$a \leqq x \leqq b,$$

and linearly independent in this interval. Let

$$\phi(x) = c_1 p_1(x) + c_2 p_2(x) + \cdots + c_n p_n(x)$$

be an arbitrary linear combination of these functions with constant coefficients, and let H be the maximum of $|\phi(x)|$ in (a, b). Then the following lemma holds:†

Lemma I. *There exists a constant Q, completely determined by the system of functions $p_1(x), \cdots, p_n(x)$, such that*

$$|c_k| \leqq QH \qquad (k = 1, 2, \cdots, n),$$

for all functions‡ $\phi(x)$.

For each value of k, let the coefficients in the expression

$$\Phi_k(x) = c_{1k} p_1(x) + c_{2k} p_2(x) + \cdots + c_{nk} p_n(x)$$

be determined so that

$$\int_a^b p_i(x)\Phi_k(x)\,dx = 0, \qquad i \neq k; \qquad \int_a^b p_k(x)\Phi_k(x)\,dx = 1.$$

* Cf., e.g., Sibirani, *Sulla rappresentazione approssimata delle funzioni*, Annali di matematica pura ed applicata, ser. 3, vol. 16 (1909), pp. 203–221.

† This lemma is given, with a somewhat different proof, by Sibirani, loc. cit., p. 208. For the polynomial case, a variety of demonstrations have been given: see Kirchberger, loc. cit., pp. 7–9; Borel, loc. cit., pp. 83–84; Tonelli, *I polinomi d' approssimazione di Tchebychev*, Annali di matematica pura ed applicata, ser. 3, vol. 15 (1908), pp. 47–119; pp. 61–62; cf. also Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, vol. I, pp. 374–375, and, for bibliographical references, vol. II, p. 896.

‡ It is evident that the statement is not true if $p_1(x), \cdots, p_n(x)$ are linearly dependent, for, if there is a combination $\phi(x)$, with coefficients not all zero, which vanishes identically, this can be multiplied by a constant so as to give a combination which has arbitrarily large coefficients, and is still identically zero; and this can be added to a combination for which $H \neq 0$ so as to contradict the lemma.

This amounts to subjecting the n coefficients $c_{1k}, \cdots, c_{nk}$ to a set of n simultaneous linear equations. The determinant of the equations is not zero,* for if it were, a set of coefficients, not all zero, could be determined for a function

$$\Phi_0(x) = c_{10}\, p_1(x) + c_{20}\, p_2(x) + \cdots + c_{n0}\, p_n(x)$$

so as to make

$$\int_a^b p_i(x)\Phi_0(x)\,dx = 0 \qquad (i = 1, 2, \cdots, n).$$

It would follow from the last set of equations, however, that

$$\int_a^b [\Phi_0(x)]^2\,dx = \int_a^b [c_{10}\, p_1(x) + \cdots + c_{n0}\, p_n(x)]\Phi_0(x)\,dx = 0,$$

and this would imply that $\Phi_0(x) = 0$ identically, which is impossible, since $p_1(x), \cdots, p_n(x)$ are linearly independent. It is certain, therefore, that the desired functions $\Phi_k(x)$ exist.

Let Q' be the greatest value attained by the absolute value of any $\Phi_k(x)$ in (a, b). Then

$$\left|\int_a^b \phi(x)\Phi_k(x)\,dx\right| \leqq Q' H(b-a).$$

On the other hand, from the definition of $\Phi_k(x)$,

$$\int_a^b \phi(x)\Phi_k(x)\,dx = c_k.$$

Consequently, if $Q = Q'(b-a)$,

$$|c_k| \leqq QH.$$

3. **Second lemma on bounds of coefficients.** Let m be a fixed number greater than 1 (not necessarily an integer). Let

$$\Delta_m = \int_a^b |\phi(x)|^m\,dx,$$

and let $\Delta = b - a$. The following lemma is analogous to that already proved:

LEMMA II. *There exists a constant* Q_1, *completely determined by the system of functions* $p_1(x), \cdots, p_n(x)$, *and, in particular, independent of* m, *such that*

$$|c_k| \leqq Q_1(\Delta + \Delta_m) \qquad (k = 1, 2, \cdots, n),$$

for all functions $\phi(x)$.

In the first place, since $m > 1$, $|\phi(x)| \leqq |\phi(x)|^m$ unless $|\phi(x)| < 1$, so

* The non-vanishing of this Gramian determinant is a well-known condition for linear independence; cf., e.g., Kowalewski, *Einführung in die Determinantentheorie*, pp. 320–325.

that, in any case,

$$|\phi(x)| \leqq 1 + |\phi(x)|^m .$$

Hence

$$\int_a^b |\phi(x)| dx \leqq \Delta + \Delta_m ,$$

and for any value of x between a and b,

$$\left| \int_a^x \phi(x) dx \right| \leqq \int_a^x |\phi(x)| dx \leqq \Delta + \Delta_m . \tag{1}$$

On the other hand, the n functions

$$\int_a^x p_1(x) dx , \int_a^x p_2(x) dx , \cdots , \int_a^x p_n(x) dx \tag{2}$$

are linearly independent, since a linear relation between them would give by differentiation a linear relation connecting $p_1(x), \cdots, p_n(x)$. Therefore, if Q_1 is the constant of Lemma I for the functions (2), it can be inferred from (1), that is, from

$$\left| c_1 \int_a^x p_1(x) dx + \cdots + c_n \int_a^x p_n(x) dx \right| \leqq \Delta + \Delta_m ,$$

that

$$|c_k| \leqq Q_1 (\Delta + \Delta_m) .$$

4. **Third lemma on bounds of coefficients.** In addition to the notation of the preceding sections, let $f(x)$ be a function continuous for $a \leqq x \leqq b$, arbitrary at the outset, but to be kept unchanged throughout the remainder of the discussion; let M be the maximum of $|f(x)|$ in (a, b); and let

$$\delta_m = \int_a^b |f(x) - \phi(x)|^m dx .$$

A further development of the ideas of the first two lemmas leads to the following statement:

LEMMA III. *For all functions* $\phi(x)$,

$$|c_k| \leqq Q_1 (M\Delta + \Delta + \delta_m) \qquad (k = 1, 2, \cdots, n),$$

where Q_1 *is the constant of the preceding lemma.*

By an appropriate modification of a remark made in the preceding section, it is recognized that

$$|f(x) - \phi(x)| \leqq 1 + |f(x) - \phi(x)|^m .$$

Hence

$$|\phi(x)| \leqq M + 1 + |f(x) - \phi(x)|^m ,$$

and

$$\int_a^b |\phi(x)| dx \leqq (M + 1)\Delta + \delta_m .$$

The concluding steps of § 3, applied to the present case, show that

$$|c_k| \leqq Q_1[(M+1)\Delta + \delta_m].$$

5. **Existence of an approximating function for exponent** m. If the function $f(x)$, the system $p_1, \cdots, p_n$, and the exponent m are given, and the coefficients c_k are regarded as undetermined, the value of δ_m, which is a function of these coefficients, has a lower limit γ_m which is positive or zero. If there is a function $\phi(x)$ for which δ_m actually attains its lower limit, this $\phi(x)$ will be called, for brevity, an *approximating function for the exponent* m. It is readily deduced from Lemma III that such a function will always exist.* For sets of coefficients $c_k^{(j)}$ can be chosen successively, $j = 1, 2, \cdots$, so that, if $\delta_m^{(j)}$ is the corresponding value of δ_m in each case,

$$\lim_{j=\infty} \delta_m^{(j)} = \gamma_m.$$

If $c_1^{(j)}, c_2^{(j)}, \cdots, c_n^{(j)}$ are regarded as the coördinates of a point P_j in space of n dimensions, all the points P_j from a certain value of j on, as soon as $\delta_m^{(j)}$ becomes and remains less than $\gamma_m + 1$, say, will lie in a bounded region,

$$|c_k^{(j)}| \leqq Q_1(M\Delta + \Delta + \gamma_m + 1).$$

The points P_j will have a limit point P in this region, and as the dependence of δ_m on the c's is continuous, the function $\phi(x)$ formed with the coefficients corresponding to the point P will make δ_m equal to γ_m. This approximating function $\phi(x)$ will be denoted by $\phi_m(x)$.

6. **Uniqueness of the approximating function for exponent** m. For each value of m, the approximating function $\phi_m(x)$ is unique. Suppose there were two such functions, $\phi_{\mathrm{I}}(x)$ and $\phi_{\mathrm{II}}(x)$, the subscript m being understood. Let

$$\phi_{\mathrm{III}}(x) = \tfrac{1}{2}[\phi_{\mathrm{I}}(x) + \phi_{\mathrm{II}}(x)],$$

and let $\delta_{\mathrm{I}}, \delta_{\mathrm{II}}, \delta_{\mathrm{III}}$, be the corresponding values of δ_m, so that $\delta_{\mathrm{I}} = \delta_{\mathrm{II}} = \gamma_m$. Furthermore, let

$$r_{\mathrm{I}}(x) = f(x) - \phi_{\mathrm{I}}(x), \qquad r_{\mathrm{II}}(x) = f(x) - \phi_{\mathrm{II}}(x),$$

$$r_{\mathrm{III}}(x) = f(x) - \phi_{\mathrm{III}}(x).$$

Then

$$r_{\mathrm{III}}(x) = \tfrac{1}{2}[r_{\mathrm{I}}(x) + r_{\mathrm{II}}(x)].$$

Since $m > 1$,

$$|r_{\mathrm{III}}(x)|^m \leqq \tfrac{1}{2}|r_{\mathrm{I}}(x)|^m + \tfrac{1}{2}|r_{\mathrm{II}}(x)|^m; \tag{3}$$

if $r_{\mathrm{I}}(x) = X_1$, for any particular value of x, $r_{\mathrm{II}}(x) = X_2$, and $r_{\mathrm{III}}(x) = X_3$, the assertion is that

$$\left|\frac{X_1 + X_2}{2}\right|^m \leqq \frac{|X_1|^m + |X_2|^m}{2},$$

* Cf. Young, loc. cit., p. 335.

which is a consequence of the fact that the graph of the function $Y = |X|^m$ is concave upward.* Moreover, the sign of inequality holds in (3), for any value of x for which $r_{\mathrm{I}} \neq r_{\mathrm{II}}$, that is, whenever $\phi_{\mathrm{I}} \neq \phi_{\mathrm{II}}$. Therefore, if ϕ_{I} and ϕ_{II} are not identically equal,

$$\int_a^b |r_{\mathrm{III}}(x)|^m\, dx < \frac{1}{2}\int_a^b |r_{\mathrm{I}}(x)|^m\, dx + \frac{1}{2}\int_a^b |r_{\mathrm{II}}(x)|^m\, dx,$$

since the integrands are continuous, and the relation (3) is an inequality over a part at least of the interval of integration. That is,

$$\delta_{\mathrm{III}} < \tfrac{1}{2}(\delta_{\mathrm{I}} + \delta_{\mathrm{II}}),$$

or, since $\delta_{\mathrm{I}} = \delta_{\mathrm{II}} = \gamma_m$,

$$\delta_{\mathrm{III}} < \gamma_m.$$

This is inconsistent with the definition of γ_m as the least possible value of δ_m.

Similar reasoning shows that no function $\phi(x)$, other than $\phi_m(x)$, can give even a relative minimum for δ_m as a function of $c_1, \cdots, c_n$. Let $\phi_{\mathrm{II}}(x)$ be any such function $\phi(x)$, let $\phi_{\mathrm{I}}(x) = \phi_m(x)$, and let

$$\phi_{\mathrm{III}}(x) = A\phi_{\mathrm{I}}(x) + B\phi_{\mathrm{II}}(x),$$

where A and B are any two positive constants whose sum is 1. Let $r_{\mathrm{I}}(x)$, $r_{\mathrm{II}}(x)$, $r_{\mathrm{III}}(x)$, and $\delta_{\mathrm{I}}, \delta_{\mathrm{II}}, \delta_{\mathrm{III}}$, be the corresponding values of $f(x) - \phi(x)$ and of δ_m. Then

$$|r_{\mathrm{III}}(x)|^m \leqq A|r_{\mathrm{I}}(x)|^m + B|r_{\mathrm{II}}(x)|^m,$$

the inequality holding whenever $\phi_{\mathrm{I}} \neq \phi_{\mathrm{II}}$. Consequently

$$\delta_{\mathrm{III}} < A\delta_{\mathrm{I}} + B\delta_{\mathrm{II}}, \tag{4}$$

or, since $\delta_{\mathrm{I}} < \delta_{\mathrm{II}}$ and $A + B = 1$,

$$\delta_{\mathrm{III}} < \delta_{\mathrm{II}}.$$

This means, inasmuch as A can be taken arbitrarily small and B arbitrarily near to 1, that it is possible to find functions $\phi(x)$ with coefficients as close to those of $\phi_{\mathrm{II}}(x)$ as may be desired, so that $\delta_m < \delta_{\mathrm{II}}$.

The main conclusions obtained hitherto (not including the last one) can be summarized as follows:

THEOREM I. *For each value of $m > 1$, there exists one and just one approximating function $\phi_m(x)$.*

7. Necessary and sufficient condition† for the approximating function $\phi_m(x)$. Let $\phi_m(x)$ be the approximating function for exponent m as before, and let

$$r_m(x) = f(x) - \phi_m(x).$$

* Analytically, of course, an immediate proof is obtained from the mean value theorem and the fact that dY/dX is an increasing function of X.

† This section is inserted for its own sake, and is not needed for what follows.

When $r_m(x) \neq 0$, let $r_m^{[m-1]}(x)$ be used as an abbreviation for the expression $|r_m(x)|^m/[r_m(x)]$, and let $r_m^{[m-1]}(x) = 0$ when $r_m(x) = 0$ (it is assumed throughout that $m > 1$). Then $r_m^{[m-1]}(x)$ is a quantity having its absolute value equal to $|r_m(x)|^{m-1}$, and having the same algebraic sign as $r_m(x)$ itself; if m is an even integer, $r_m^{[m-1]}(x)$ is simply $[r_m(x)]^{m-1}$. It will be shown that $r_m^{[m-1]}(x)$ must be orthogonal to each of the functions $p_k(x)$ in the interval (a, b):

$$\int_a^b r_m^{[m-1]}(x)\,p_k(x)\,dx = 0 \qquad (k = 1, 2, \cdots, n).$$

To bring out what is essential in the proof, let it be given first for the special case $m = 2$. Let $p(x)$, without subscript, stand for any one of the functions $p_k(x)$, and let h be an arbitrary constant, positive, negative, or zero. Let

$$\phi(x) = \phi_2(x) + hp(x),$$

$$r(x) = f(x) - \phi(x) = r_2(x) - hp(x);$$

then

$$|r(x)|^2 = [r(x)]^2 = [r_2(x)]^2 - 2hr_2(x)\,p(x) + h^2[p(x)]^2.$$

Hence

$$\delta_2 = \int_a^b |r(x)|^2\,dx = \gamma_2 - 2h\int_a^b r_2(x)\,p(x)\,dx + h^2\int_a^b [p(x)]^2\,dx,$$

since $r_2(x)$ is understood to be the error of the approximating function for $m = 2$, so that

$$\int_a^b [r_2(x)]^2\,dx = \gamma_2.$$

In the relation

$$\delta_2 = \gamma_2 - h\left[2\int_a^b r_2(x)\,p(x)\,dx - h\int_a^b [p(x)]^2\,dx\right],$$

suppose that the first of the two terms of the expression in brackets is not zero; it is to be shown that this leads to a contradiction. If h is sufficiently small numerically, the second term will be smaller numerically than the first, and the value of the whole bracket will be different from zero and will have the sign of the first term. If h is given a small value of the same sign as the first term in the bracket, the whole expression to be subtracted from γ_2 will be positive, and the value of δ_2 corresponding to the function $\phi(x)$ will be smaller than γ_2. Since this is contrary to the definition of γ_2, the truth of the assertion is established in the special case.

It is evident that an altogether similar proof can be given if m is any even integer. The demonstration can be modified so as to make it applicable to other cases as well. In general, let

$$\phi(x) = \phi_m(x) + hp(x),$$

$$r(x) = f(x) - \phi(x) = r_m(x) - hp(x),$$

with the understanding that

$$\delta_m = \int_a^b |r(x)|^m\,dx, \qquad \gamma_m = \int_a^b |r_m(x)|^m\,dx.$$

Then

$$\frac{d}{dh}\delta_m = \int_a^b \frac{\partial}{\partial h}|r(x)|^m\,dx.$$

If $r(x) > 0$,

$$\frac{\partial}{\partial h}|r(x)|^m = \frac{\partial}{\partial h}[r(x)]^m = m[r(x)]^{m-1}\frac{\partial}{\partial h}[r(x)]$$
$$= -mp(x)[r(x)]^{m-1} = -mp(x)|r(x)|^{m-1}.$$

If $r(x) < 0$,

$$\frac{\partial}{\partial h}|r(x)|^m = \frac{\partial}{\partial h}[-r(x)]^m = m[-r(x)]^{m-1}\frac{\partial}{\partial h}[-r(x)]$$
$$= mp(x)[-r(x)]^{m-1} = mp(x)|r(x)|^{m-1}.$$

If $r(x) = 0$,

$$\frac{\partial}{\partial h}|r(x)|^m = 0,$$

whether h is given positive or negative increments. In any case,

$$\frac{\partial}{\partial h}|r(x)|^m = -mp(x)|r(x)|^m/[r(x)],$$

a continuous function of x and h, the value of the fraction being taken to be zero when $r(x) = 0$, and

$$\left[\frac{\partial}{\partial h}|r(x)|^m\right]_{h=0} = -mp(x)r_m^{[m-1]}(x),$$

so that

$$\left[\frac{d}{dh}\delta_m\right]_{h=0} = -m\int_a^b p(x)r_m^{[m-1]}(x)\,dx.$$

The last integral must be zero, otherwise it would be possible to give h a small value, positive or negative, so as to make

$$\delta_m(h) < \delta_m(0),$$

that is,

$$\delta_m(h) < \gamma_m,$$

which is inadmissible. So the assertion made at the beginning of the section is true in general.

It is merely another statement of the same conclusion to say that $r_m^{[m-1]}(x)$ must be orthogonal to every function $\phi(x)$.

The necessary condition that has been obtained for $\phi_m(x)$ is also sufficient. This follows from the reasoning in the latter part of § 6, which led up to the

remark that no function $\phi(x)$, other than $\phi_m(x)$, can give even a relative minimum for δ_m. Suppose that $\phi_{\mathrm{I}}(x)$ is a linear combination of the functions $p_k(x)$, not identical with $\phi_m(x)$. Let

$$r_{\mathrm{I}}(x) = f(x) - \phi_{\mathrm{I}}(x),$$

and let $r_{\mathrm{I}}^{[m-1]}(x)$ be defined in a manner corresponding to the definition of $r_m^{[m-1]}(x)$ above. It is to be shown that there exists a function $\psi(x)$ which is a linear combination of the p's, such that

$$\int_a^b \psi(x) r_{\mathrm{I}}^{[m-1]}(x)\, dx \neq 0. \tag{5}$$

For any linear combination $\psi(x)$, let

$$\phi(x) = \phi_{\mathrm{I}}(x) + h\psi(x),$$

$$r(x) = f(x) - \phi(x) = r_{\mathrm{I}}(x) - h\psi(x),$$

$$\delta_m = \int_a^b |r(x)|^m\, dx, \qquad \delta_{\mathrm{I}} = \int_a^b |r_{\mathrm{I}}(x)|^m\, dx.$$

By a calculation corresponding to that of the third paragraph preceding, it is seen that

$$\left[\frac{d}{dh}\delta_m\right]_{h=0} = -m\int_a^b \psi(x) r_{\mathrm{I}}^{[m-1]}(x)\, dx. \tag{6}$$

Now let

$$\psi(x) = \phi_m(x) - \phi_{\mathrm{I}}(x);$$

then

$$\phi(x) = h\phi_m(x) + (1-h)\phi_{\mathrm{I}}(x).$$

For positive values of h, the inequality (4) of § 6 is applicable with A, B, δ_{III}, δ_{I}, and δ_{II} replaced by h, $1-h$, δ_m, γ_m, and δ_{I} respectively:

$$\delta_m < h\gamma_m + (1-h)\delta_{\mathrm{I}},$$

that is,

$$\delta_m < \delta_{\mathrm{I}} + h(\gamma_m - \delta_{\mathrm{I}}), \qquad \frac{\delta_m - \delta_{\mathrm{I}}}{h} < \gamma_m - \delta_{\mathrm{I}},$$

the difference $\gamma_m - \delta_{\mathrm{I}}$ being negative. Consequently

$$\left[\frac{d}{dh}\delta_m\right]_{h=0} \leqq \gamma_m - \delta_{\mathrm{I}} < 0,$$

and, because of (6), the inequality (5) is verified.

To summarize, using the symbols $r(x)$ and $r^{[m-1]}(x)$ in a manner corresponding to the previous notation:*

THEOREM II. *In order that $\phi(x)$ be the approximating function for exponent*

* That is, $r(x) = f(x) - \phi(x)$, $r^{[m-1]}(x) = |r(x)|^m/[r(x)]$ when $r(x) \neq 0$, $r^{[m-1]}(x) = 0$ when $r(x) = 0$.

m, it is necessary and sufficient that

$$\int_a^b \psi(x)\, r^{[m-1]}(x)\, dx = 0$$

for all functions $\psi(x)$ which are linear combinations of $p_1(x), \cdots, p_n(x)$.

If the functions $p_k(x)$ are the quantities x^{k-1}, $k = 1, 2, \cdots, n$, it can be inferred further that $r(x)$, if not identically zero, *must change sign at least n times in the interval (a, b)*, for any value of m. Otherwise it would be possible to assign ν points $x_1, x_2, \cdots, x_\nu$, $\nu \leqq n - 1$, so that $r(x)$, and hence $r^{[m-1]}(x)$, would be of constant sign (wherever different from zero) in each of the intervals $a \leqq x \leqq x_1$, $x_1 \leqq x \leqq x_2$, $\cdots$, $x_\nu \leqq x \leqq b$, and would take on opposite signs in successive intervals. Then the polynomial*

$$\psi(x) = (x - x_1)(x - x_2) \cdots (x - x_\nu),$$

of degree $\leqq n - 1$, would certainly not be orthogonal to $r^{[m-1]}(x)$, since the product $\psi(x)\, r^{[m-1]}(x)$, continuous and not vanishing identically, would be of constant sign wherever different from zero. Similar reasoning is possible in a class of other cases, including that of approximation by finite trigonometric sums, but of course not in the case of arbitrary functions $p_k(x)$.

8. **Limit of maximum error of $\phi_m(x)$ as m becomes infinite.** In this section and the following one, it will be assumed for convenience that $|f(x)| < 1$ for $a \leqq x \leqq b$. It will turn out that this is no real restriction of generality for the main conclusions, since multiplication of $f(x)$ by any constant corresponds to multiplication of the approximating functions $\phi_m(x)$, and of the other approximating functions to be considered, by the same constant.

For any function $\phi(x)$ (that is, any linear combination of the p's) let l be the maximum value of $|f(x) - \phi(x)|$ in (a, b); let l_m be the maximum of $|f(x) - \phi_m(x)|$, and let l_0 be the lower limit of l for all possible functions $\phi(x)$. It can be inferred from Lemma I that there is at least one $\phi(x)$ for which the limit l_0 is attained.† For l is a continuous function of the coefficients of ϕ, and all the coefficients of any combination ϕ for which l is near l_0 belong to a restricted region in the space of $c_1, \cdots, c_n$, so that there will be some set of values of these parameters for which l reaches its limit. Let the function ϕ corresponding to such a set of coefficients be denoted by‡ $\phi_0(x)$. It may be spoken of as the Tchebychef function, or a Tchebychef function, for $f(x)$; the question of its uniqueness need not be raised until the following section. The purpose of the present section is to prove:

* If $r(x)$ did not change sign at all, it would be understood that $\psi(x) = 1$.

† Cf. Young, loc. cit., p. 335; Fréchet, Annales de l'École Normale Supérieure, loc. cit., p. 45; Sibirani, loc. cit., p. 210.

‡ In view of what follows, it would be more suggestive to represent this function by $\phi_\infty(x)$, and the corresponding maximum error by l_∞, but it is not necessary to anticipate to that extent.

THEOREM III. *As m becomes infinite, l_m approaches the limit l_0.*

From the hypothesis that $|f(x)| < 1$, it follows that γ_m, the lower limit of δ_m for all possible functions $\phi(x)$, is less than $b - a$, for all values of m. For the particular function $\phi(x) = 0$ makes

$$\delta_m = \int_a^b |f(x)|^m \, dx < b - a,$$

and γ_m must be less than or equal to this δ_m. Hence, if c_k is any coefficient of any function $\phi_m(x)$, the term δ_m in the inequality of Lemma III may be replaced by $\Delta = b - a$, while $M < 1$, so that

$$|c_k| \leqq 3Q_1 \Delta. \tag{7}$$

That is, the absolute values of the coefficients have an upper bound which is independent of m.

Let ϵ be any positive quantity, and suppose that $|f - \phi_m| \geqq l_0 + \epsilon$ for some value $x = x_0$ in (a, b):

$$|f(x_0) - \phi_m(x_0)| \geqq l_0 + \epsilon. \tag{8}$$

Since $f(x)$ is continuous for $a \leqq x \leqq b$, it is uniformly continuous there. Let δ' be a positive quantity such that

$$|f(x') - f(x'')| \leqq \tfrac{1}{3}\epsilon$$

for $|x' - x''| \leqq \delta'$; in particular,

$$|f(x) - f(x_0)| \leqq \tfrac{1}{3}\epsilon \tag{9}$$

for $|x - x_0| \leqq \delta'$. Each of the functions $p_k(x)$ is likewise uniformly continuous in (a, b); let δ'' be so small that

$$|p_k(x') - p_k(x'')| \leqq \frac{\epsilon}{9nQ_1 \Delta}$$

whenever $|x' - x''| \leqq \delta''$, for all values of k. In view of (7) and the fact that there are n terms in $\phi_m(x)$, it follows that

$$|\phi_m(x) - \phi_m(x_0)| \leqq \tfrac{1}{3}\epsilon \tag{10}$$

if $|x - x_0| \leqq \delta''$. Let δ be the smaller of the quantities δ', δ''; then, as a consequence of (8), (9), and (10),

$$|f(x) - \phi_m(x)| \geqq l_0 + \tfrac{1}{3}\epsilon, \tag{11}$$

for $|x - x_0| \leqq \delta$, where δ is independent of m, though different values of m may call for different values of x_0. If it be supposed further that $\delta < \frac{1}{2}(b - a)$, then at least one of the intervals $(x_0 - \delta, x_0)$, $(x_0, x_0 + \delta)$ will be wholly

contained in (a, b), wherever x_0 may be, and there will certainly be an interval of length δ at least throughout which (11) is satisfied.

Then

$$\int_a^b |f(x) - \phi_m(x)|^m \, dx \geqq (l_0 + \tfrac{1}{3}\epsilon)^m \delta. \tag{12}$$

On the other hand,

$$\int_a^b |f(x) - \phi_0(x)|^m \, dx \leqq l_0^m (b - a). \tag{13}$$

But m can be taken so large as to make the right-hand member of (12) larger than the right-hand member of (13). If (12) were still to hold, $\phi_0(x)$ would give a smaller value of δ_m than $\phi_m(x)$, which would be inconsistent with the definition of $\phi_m(x)$ as the function giving the smallest possible value of δ_m. So (12), and with it the hypothesis on which (12) is based, namely the inequality (8), must cease to be true. That is, for all values of m from a certain point on,

$$|f(x) - \phi_m(x)| < l_0 + \epsilon$$

throughout (a, b), and this is equivalent to the assertion of Theorem III.

9. **Limit of $\phi_m(x)$ as m becomes infinite.** In consequence of (7), the coefficients c_k of $\phi_m(x)$, regarded as coördinates of a point in space of n dimensions, must give rise to at least one limit point as m becomes infinite. From Theorem III, with the fact that l, the maximum of $|f(x) - \phi(x)|$, is a continuous function of the coefficients of ϕ, it follows that the value of l for any function ϕ corresponding to such a limiting set of coefficients must be l_0. It is known, however, that in the case of approximation by polynomials* or by finite trigonometric sums,† and in an extensive class of cases generally,‡ there can be only one function $\phi(x)$ for which the limit l_0 is attained. For these circumstances, the statement of Theorem III can be given the more striking form:

THEOREM IV. *If the system of functions $p_k(x)$ is such that the Tchebychef function $\phi_0(x)$ is uniquely determined, then*

$$\lim_{m=\infty} \phi_m(x) = \phi_0(x),$$

in the sense that the coefficients of ϕ_m approach those of ϕ_0, and the value of $\phi_m(x)$ therefore approaches that of $\phi_0(x)$ uniformly for $a \leqq x \leqq b$.

UNIVERSITY OF MINNESOTA,
MINNEAPOLIS, MINN.

* Cf. Kirchberger, Borel, locc. citt.

† Cf. Young, Fréchet, Tonelli, locc. citt.

‡ Cf. Young, Fréchet, Sibirani, locc. citt.